An Experimental 1-km Warn-on-Forecast System for Hazardous Weather Events

Yaping Wang[1*1,2], Nusrat Yussouf[1,2,3], Christopher A. Kerr[1,2], Derek R. Stratman[1,2] and Brian C. Matilla[1,2]

[1]*Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*

[2]*NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*

[3]*School of Meteorology, University of Oklahoma, Norman, Oklahoma*

[1*] *Corresponding author:* Yaping Wang, (yaping.wang@noaa.gov), 120 David L. Boren Blvd., Norman, OK 73072

Early Online Release: This preliminary version has been accepted for publication in *Monthly Weather Review* cited, and has been assigned DOI 10.1175/MWR-D-22-0094.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

An experimental Warn-on-Forecast (WoF) ensemble data assimilation (DA) and prediction system at 1-km grid spacing is developed and tested using two landfalling tropical cyclone (TC) events, one springtime severe thunderstorm event, and one summertime flash flood event. To evaluate the impact of DA at 1-km grid spacing, two experiments are conducted. One experiment, namely the WoFS-1km, generates 3-h ensemble forecasts from the 1-km WoFS analyses while another experiment, namely the Downscaled-1km, generates 3-h ensemble forecasts from downscaled 3-km analyses. With 1-km DA, the two landfalling TC events and the summertime event show some improvement in predicting high reflectivity, while the springtime event performs worse. Meanwhile, WoFS-1km is slightly better at predicting heavier precipitation (>20 mm h^{-1}) with lower bias. However, heavy precipitation spatial placement error is only mitigated in one TC event and the summertime event with 1-km DA but is neutral or worse in the other two events. Object-based verification for rotation objects indicates that WoFS-1km performs better in one of the TC events, but worse in the springtime event with lower probability of detection and higher false alarm ratio due to fewer strong rotation objects being generated. The forecast skill of WoFS-1km for the springtime event is degraded mainly because the convective cores do not sufficiently develop as the forecast advances. The conditional benefits from 1-km DA in this study highlights the need for evaluation of a larger sample of convective storm cases and further development of the system.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

1. Introduction

The National Severe Storms Laboratory's (NSSL) experimental Warn-on-Forecast System (WoFS) is a storm-scale frequent update-cycle ensemble data assimilation (DA) and prediction system designed to provide continuous 0–6-h probabilistic guidance of severe storm hazards (e.g., tornadoes, flash flooding, large hail, severe winds, and lightning rate). The baseline WoFS is developed using a 3-km horizontal grid spacing (WoFS-3km) and has demonstrated the capability to provide skillful short-term probabilistic forecasts of thunderstorm hazards (Wheatley et al. 2015; Yussouf et al. 2013, 2015; Jones et al. 2016, 2018; Skinner et al. 2018) and heavy rainfall events (Lawson et al. 2018; Yussouf et al. 2016, 2020; Yussouf and Knopfmeier 2019; Jones et al. 2019). Since 2017, the WoFS-3km has been running as an on-demand system during spring and summer experiments for testing and evaluation by the operational forecasters at NOAA national centers and several local offices. The system is showing promise in influencing the lead-time, specificity, and uncertainty information of storm hazards at the "Watch-to-Warning" scale (Wilson et al. 2019a, b).

Presently, operational storm-scale numerical weather prediction (NWP) models in the United States typically run with a 3-km horizontal grid spacing (Δx hereafter), such as the High-Resolution Rapid Refresh (HRRR) model (Benjamin et al. 2016) or the High-Resolution Ensemble Forecast (HREF) system (Roberts et al. 2019). However, convection-allowing models (CAMs) with 3–4-km Δx have practical predictability limits on severe convective hazards owing to model resolution error. To fully resolve deep moist convection, Δx of ~100 m or less (Bryan et al. 2003) is needed.

Decreasing Δx to below ~3 km in CAMs has been explored by many studies (e.g., Kain et al. 2008; Schwartz et al. 2009, 2014; Johnson et al. 2013; Loken et al. 2017; Xue et al. 2013). Recently, Schwartz et al. (2017) found that 1-km ensemble probabilistic forecasts were especially better over the first 12 hours and at heavier rainfall rates (i.e., $\geq$ 5 mm h^{-1}) for the central and eastern portions of the CONUS when compared to the 3-km ensemble forecasts. They attributed this improvement to more accurate placement of mesoscale convective systems in the higher-resolution simulations. The benefit of decreasing Δx to improve the storm motion forecast was also demonstrated by VandenBerg et al. (2014), which compares the storm motion from 1-km and 4-km Δx forecasts. However, Schumacher (2015) found that the higher resolution simulation fails to initiate a supercell and alters the evolution of the subsequent convection. Schwartz and Sobash

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

(2019) found that 1-km model climatologies of precipitation aligned better with observations than 3-km climatologies using forecasts from deterministic convection-allowing NWP models. They pointed out that the 1-km forecasts were more skillful than the 3-km forecasts during winter and spring while having similar skill during summer. Sobash et al. (2019) demonstrated that the next-day tornado surrogate severe probability forecasts produced from the 1-km Δx CAM forecasts were more skillful than the 3-km forecasts when verified against tornado reports and warnings, indicating that the development of near-surface rotation could be better captured as Δx decreases. The results from many of the previous studies are encouraging, indicating promising benefits of higher-resolution forecast systems.

There have been several recent studies examining the potential benefits of finer-grid spacing (Δx≈1 km) forecasts in the WoF paradigm. In an idealized framework, Potvin and Flora (2015) examined the sensitivity of supercell simulations to Δx, suggesting that decreasing Δx from 3 km to 1 km produces useful rapid changes in low-level rotation intensity. In real data cases, Britt et al. (2020) found that the WoFS forecasts with a 1-km Δx may better predict the presence and cycling frequency of cyclic supercells, which is potentially useful guidance for tornadogenesis. This result is generally consistent with the findings of Adlerman and Droegemeier (2002) that distinct mesocyclone cycles were reproduced by decreasing model Δx from 2 km to 0.5 km. Lawson et al. (2021) found that the downscaled 1-km forecasts from the WoFS-3km are less skillful in predicting weak-moderate reflectivity but could better detect high reflectivity and rapid low- and midlevel rotations than the WoFS-3km forecast, even though the improvements were case-sensitive. Miller et al. (2021) found that downscaling the WoFS-3km forecasts to 1.5 km does not improve reflectivity object occurrence but performs better in detecting midlevel mesocyclones through both deterministic and probabilistic verification approaches.

Nevertheless, the higher-resolution forecasts in these studies are all initialized from the downscaled coarser-resolution initial conditions (ICs), in which the resolution error and the unresolved finer-scale characteristics could limit forecast accuracy. Potvin et al. (2017) examined the impact of initial condition (IC) resolution error on forecasts of convection by conducting sensitivity experiments using identical model settings but coarsened ICs. They found that the qualitative storm evolution is insensitive to the IC resolution. However, Wang and Wang (2020) compared two 500-m forecasts initialized from a 500-m analysis IC and a downscaled 2-km IC

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

and found that using higher-resolution ICs could generate more realistic storm morphology and more accurate track of modeled tornado-like vortices.

While the goal of the WoFS is to provide guidance of various weather hazards with longer forecast lead times and improved spatiotemporal accuracy, this study introduces an experimental 1-km WoF DA and prediction system, in which the 1-km analyses are generated independently by assimilating various observations, rather than downscaling ICs from the parent 3-km analyses (e.g., Britt et al. 2020; Lawson et al. 2021; Miller et al. 2021). Herein, the performance of the 1-km WoFS system is evaluated for several severe weather events, including two landfalling TCs, a summertime flood-inducing storm and a springtime severe thunderstorm. The goal is to develop an experimental 1-km WoFS and explore the hypothesis that generating higher-resolution ICs better represents extreme precipitation and tornadogenesis potential across different weather systems. This study is organized as follows: Section 2 introduces the DA and prediction system along with the 1-km forecast experiments. The evaluation and verification methods are introduced in Section 3. Section 4 presents results for 1-km DA sensitivity experiments followed by verification and performance comparison results of the two sets of 1-km forecasts. Summary and conclusions are given in Section 5.

2. Experimental design

a. Description of the DA and prediction system

The DA and prediction system in this study is configured with two nested domains at 3- and 1-km Δx, respectively. The 3-km parent domain is identical to the current baseline WoFS domain, which contains 300 × 300 grid points and 51 vertical levels. The inner 1-km domain contains 402 × 402 grid points and the same vertical levels as the parent domain. The location of the 3-km and 1-km domains are case dependent and are placed over the most likely regions for severe weather hazards. The configuration of the system is very similar to the baseline WoFS that was run in a real-time setting during the 2021 NOAA Hazardous Weather Testbed (HWT) Spring Forecast Experiment (SFE, Kain et al. 2003; Gallo et al. 2017; Clark et al. 2020). The system is comprised of 36 ensemble members for both domains (Fig. 1), and uses the Advanced Research version of the Weather Research and Forecasting (WRF-ARW) model, version 3.9.1 (Skamarock et al. 2008). The experimental High-Resolution Rapid Refresh Ensemble (HRRRE; Dowell et al. 2016) provides the lateral boundary conditions for the 3-km ensemble and the ICs for both the 3-

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

and 1-km ensembles. A multi-physics ensemble configuration with different sets of planetary boundary layer and radiation schemes (e.g., Stensrud et al. 2000; Wheatley et al. 2014) are applied to both 1-km and 3-km ensembles for DA and prediction (Table 1). The ensembles use the NSSL two-moment microphysical parameterization scheme (Mansell et al. 2010) and the Noah-MP (Niu et al. 2011; Yang et al. 2011) land surface scheme. For forecasts, the 1-km domain is nested in the 3-km domain with one-way feedback; the 3-km ensemble provides boundary conditions to the 1-km ensemble and the ICs are provided by the corresponding analysis system. The time steps are 15 s and 5 s for the 3-km and 1-km runs respectively. For brevity, the 3-km and 1-km ensemble DA and prediction systems are referred to as WoFS-3km and WoFS-1km respectively.

Table 1 Configurations of physics options for the multiphysics WRF ensemble system.

Multiphysics ensemble						
Ensemble member	HRRRE member	Land surface model	Microphysics	Planetary boundary layer	Shortwave radiation	Longwave radiation
1	1	Noah-MP	NSSL 2-moment	YSU	Dudhia	RRTM
2	2			YSU	RRTMG	RRTMG
3	3			MYJ	Dudhia	RRTM
4	4			MYJ	RRTMG	RRTMG
5	5			MYNN	Dudhia	RRTM
6	6			MYNN	RRTMG	RRTMG
7	7	Noah-MP	NSSL 2-moment	YSU	Dudhia	RRTM
8	8			YSU	RRTMG	RRTMG
9	9			MYJ	Dudhia	RRTM
10	10			MYJ	RRTMG	RRTMG
11	11			MYNN	Dudhia	RRTM
12	12			MYNN	RRTMG	RRTMG
13	13	Noah-MP	NSSL 2-moment	YSU	Dudhia	RRTM
14	14			YSU	RRTMG	RRTMG
15	15			MYJ	Dudhia	RRTM
16	16			MYJ	RRTMG	RRTMG
17	17			MYNN	Dudhia	RRTM
18	18			MYNN	RRTMG	RRTMG
19	18	Noah-MP	NSSL 2-moment	YSU	Dudhia	RRTM
20	17			YSU	RRTMG	RRTMG
21	16			MYJ	Dudhia	RRTM
22	15			MYJ	RRTMG	RRTMG
23	14			MYNN	Dudhia	RRTM
24	13			MYNN	RRTMG	RRTMG
25	12	Noah-MP	NSSL 2-moment	YSU	Dudhia	RRTM

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

26	11			YSU	RRTMG	RRTMG
27	10			MYJ	Dudhia	RRTM
28	9			MYJ	RRTMG	RRTMG
29	8			MYNN	Dudhia	RRTM
30	7			MYNN	RRTMG	RRTMG
31	6	Noah-MP	NSSL 2-moment	YSU	Dudhia	RRTM
32	5			YSU	RRTMG	RRTMG
33	4			MYJ	Dudhia	RRTM
34	3			MYJ	RRTMG	RRTMG
35	2			MYNN	Dudhia	RRTM
36	1			MYNN	RRTMG	RRTMG

The cycled DA experiments for the WoFS-1km and WoFS-3km are conducted in parallel. Available storm observations are continuously assimilated into both the 3-km and 1-km ensembles at 15-min intervals using the Community Gridpoint Statistical Interpolation (GSI; e.g., Kleist et al. 2009, Shao et al. 2016) based ensemble Kalman filter (EnKF, Houtekamer et al. 2005) DA system (GSI-EnKF; Fig. 1). The same sets of conventional, GOES-16 satellite-derived cloud water path (CWP) and GOES-16 clear sky radiance are used for both WoFS-1km and WoFS-3km. For these non-radar observations in this work, the covariance localization sets for WoFS-3km (see Table 2 in Kerr et al. 2022, manuscript submitted to *Weather and Forecasting*) are applied to WoFS-1km. However, radar observations are processed separately for the two domains. Specifically, radar reflectivity and radial velocity observations, which are derived from the Multi-Radar Multi-Sensor (MRMS) product (Smith et al. 2016), are objectively analyzed to a 5-km grid for the WoFS-3km and to a 3-km grid for the WoFS-1km. In areas where the composite reflectivity is less than 15 dBZ, radar reflectivity is set to 0 dBZ (i.e., clear-air reflectivity) and radial velocity is set to missing where the composite reflectivity is less than 20 dBZ. To suppress spurious convection, clear-air reflectivity is thinned to a 15-km grid and a 6-km grid and assimilated into the WoFS-3km and WoFS-1km, respectively. To maintain ensemble spread, the spatially and temporally varying multiplicative adaptive inflation technique is applied (Anderson 2009; Hu et al. 2019) in both systems. In this study, the radar DA configurations generally follow those used in the current baseline WoFS, yet, some parameters are tested and tuned for the WoFS-1km (see Section 4a).

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

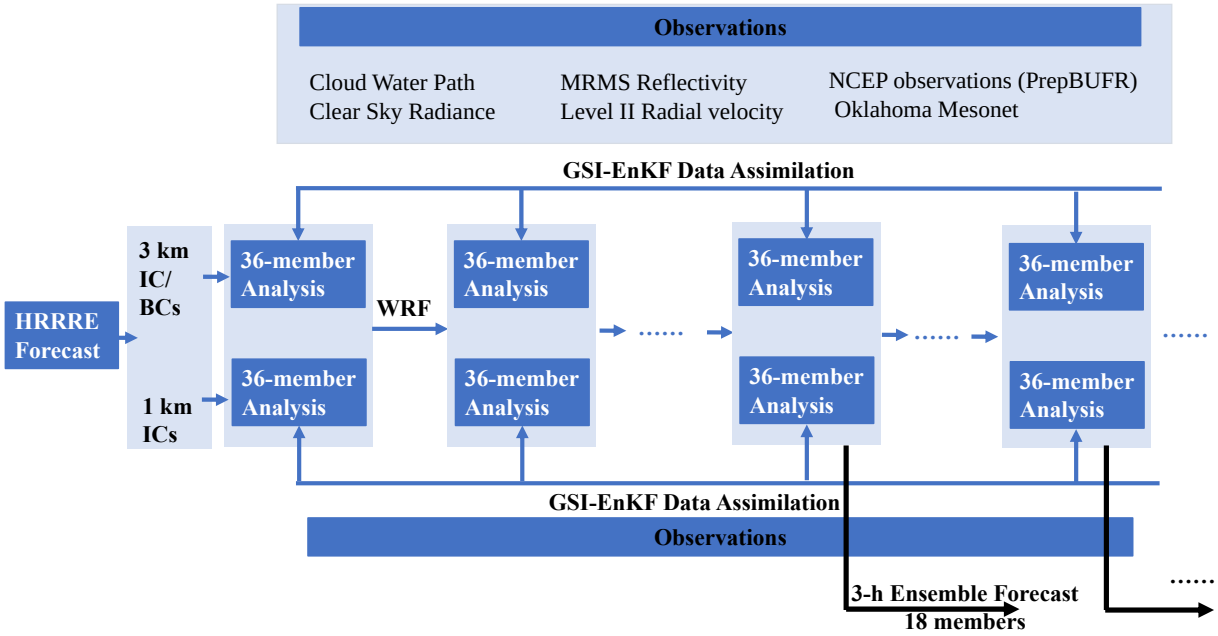


Fig.1 Flowchart of the 1- and 3-km ensemble DA and prediction system.

b. 1-km forecast experiments

As mentioned earlier, the goal of this study is to develop the 1-km WoFS and explore the potential benefit of the 1-km WoFS in providing probabilistic guidance to thunderstorm hazards. To evaluate the performance of the 1-km WoFS, four hazardous weather events are run retrospectively in this study (Table 2). The cases include two landfalling tropical cyclones (TCs), Harvey in 2017 and Michael in 2018, one summertime flash flood event in 2019, and one springtime mesoscale convective system (MCS) in 2020. The simulation domains and the locations of radar sites for the four cases are shown in Fig. 2. To evaluate the potential benefit from higher-resolution DA and to exclude any potential impact caused by different Δx, two 1-km forecast experiments within the nested domain are conducted. In one experiment, 3-h ensemble forecasts are initialized from the WoFS-1km cycled analyses, and in another experiment, 3-h ensemble forecasts are initialized at 1-km Δx by downscaling the WoFS-3km analyses. Hereafter, these experiments are referred to as WoFS-1km and Downscaled-1km, respectively. As listed in Table 2, the DA cycling for the four cases begins at different times, but all ends at 0000 UTC. Also, owing to the decreasing forecast skill after 3-h of integration and the considerable computational expense, 3-h ensemble forecasts, instead of 6-h forecasts as in the real-time experiments, are launched every hour from the first 18 members at the initialization times listed in Table 2.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Table 2 The experimental time line for the selected four cases

Case name	Date	DA cycling duration	Initialization times (UTC)
Harvey	25 Aug 2017	1800 UTC 25 Aug to 0000 UTC 26 Aug	2000, 2100, 2200, 2300, 0000
Michael	10 Oct 2018	1600 UTC 10 Oct to 0000 UTC 11 Oct	1800, 1900, 2000, 2100, 2200, 2300, 0000
Summertime	18 Jun 2019	1500 UTC 18 Jun to 0000 UTC 19 Jun	1800, 1900, 2000, 2100, 2200, 2300, 0000
Springtime	5 May 2020	1500 UTC 5 May to 0000 UTC 6 May	1800, 1900, 2000, 2100, 2200, 2300, 0000

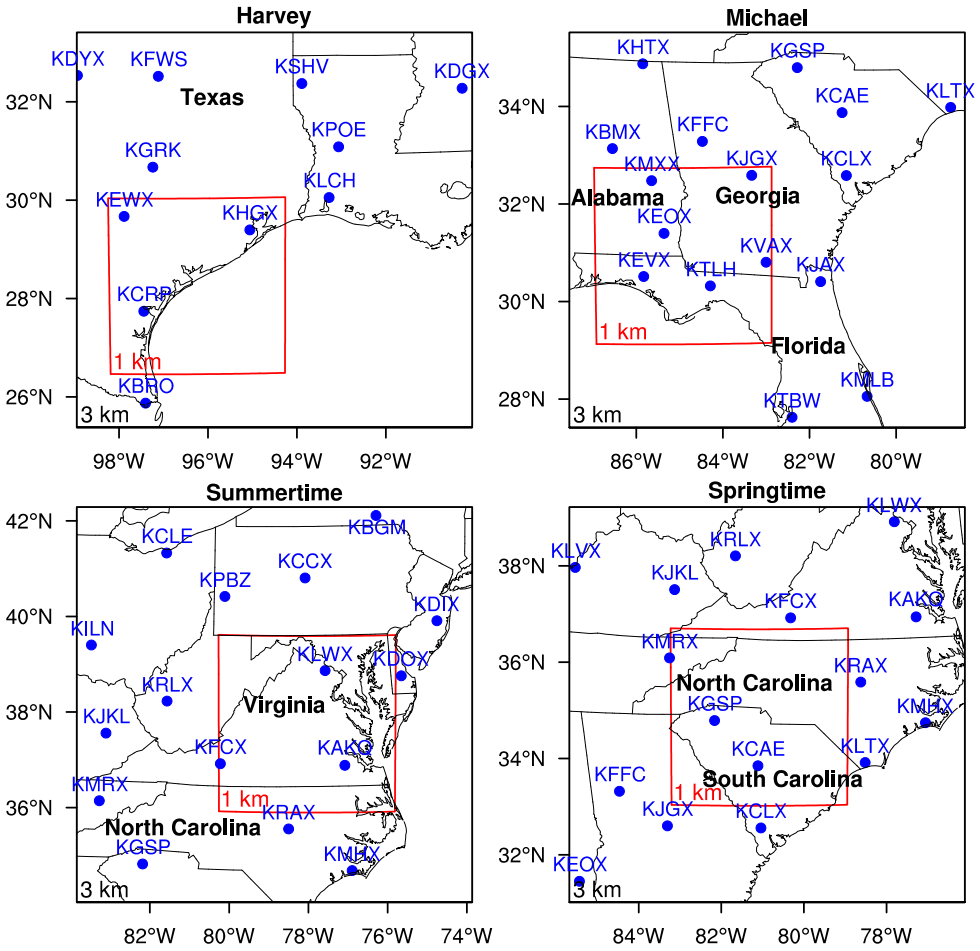


Fig. 2 The 3-km (black) and 1-km (red) domains and the locations of radar sites for the four cases.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

3. Evaluation and verification methods

To evaluate the EnKF performance of the 1-km system, observation-space diagnostics, including mean innovation, the root-mean-square innovation (RMSI), total ensemble spread, and consistency ratio (CR) are calculated using the equations described in previous studies (e.g., Dowell et al. 2004; Dowell and Wicker 2009; Dowell et al. 2011; Yussouf et al. 2013; Wheatley et al. 2015; Jones et al. 2018). The mean innovation and RMSI are the corresponding statistics between the observations and the model variables. Mean innovation is defined as the observation minus the ensemble mean prior (forecast) or posterior (analysis). It measures the model bias, with a positive value indicating an underpredicting model forecast and a negative value indicating an overpredicting forecast. The RMSI measures how the model fits the observations, so a smaller RMSI is desirable. The total spread is the square root of the summation of the observation error variance and the ensemble variance, which measures the degree of the ensemble spread. The total spread (Dowell and Wicker 2009) is defined as

$$\text{Total spread} = \sqrt{\sigma_{obs}^2 + \left\langle \frac{1}{N-1} \sum_{n=1}^{N} \left[H(\mathbf{x_n}) - \overline{H(\mathbf{x})} \right]^2 \right\rangle}, \qquad (1)$$

where σ_{obs} is the assumed observation error standard deviation, H is the forward operator that maps the model state vector to the observation space, $\mathbf{x}$ is the model state vector, and N is the ensemble size. The ratio of the square of total spread to the variance of prior innovation (square of RMSI) is defined as consistency ratio (CR). A comparable magnitude of prior total spread and RMSI, or a CR value of ~1.0 is desirable, indicating that the ensemble variance is an optimal approximation for the forecast error variance with the assumed observation error (Dowell et al. 2004).

The observed MRMS composite reflectivity, MRMS quantitative precipitation estimation (QPE) products (Zhang et al. 2016), and MRMS 0–2 km and 2–5 km rotation fields derived from MRMS azimuthal wind shear data are used in this study to assess the ensemble forecasts accuracy of reflectivity, precipitation, and low- and mid-level rotation, respectively. Only the nested domains (red boxes in Fig. 2) are used in the calculations. To assess the potential benefits of the finer-resolution system in predicting reflectivity and precipitation, ensemble fractions skill scores (eFSS; Duc et al. 2013) are computed for certain thresholds using different neighborhood radii. The eFSS extends the concept of spatial neighborhood in the traditional FSS (Roberts and Lean 2008) to the ensemble dimension (Duc et al. 2013). The application of the neighborhood approach can decrease the negative impact of small displacement errors (Schwartz et al. 2017). That is, the

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

eFSS computes the fractions of the event of interest within a certain spatial neighborhood of each grid point and averages the fractions over all the ensemble members. The data within each neighborhood but outside the domain are treated as correct negative forecasts (Nachamkin and Schmidt, 2015). The statistical significance of the eFSS difference between two experiments are assessed using a bootstrap resampling technique (e.g., Hamill 1999; Wolff et al. 2014), in which paired resamples of the forecasts from two experiments were randomly drawn and the aggregate differences were calculated. The procedure was repeated 1000 times. The differences at a 95% (5%) or higher (lower) level were determined to be statistically significant. In this work, we ignored the autocorrelation among forecasts.

The performance of WoFS-1km and Downscaled-1km experiments in forecasting potential tornado occurrence is also evaluated. Since 1-km Δx cannot resolve tornadoes, the 0–2-km and 2–5-km updraft helicity (UH) are diagnosed from the model variables and are used as proxies for tornado-producing mesocyclones in severe storms (Kain et al. 2008; Skinner et al. 2018). For the landfalling TC cases, tornado-producing mesocyclones in TCs are characterized by strong low-level rotation but little buoyancy, so the vertically-averaged vertical vorticity in the 0-2-km layer (VOR; Jones et al. 2019; Yussouf et al. 2019) is used as the proxy for rotation instead of UH. For the TC cases in this work, the results of using 0–2-km UH and 0–2-km VOR are consistent. Correspondingly, the MRMS 0–2-km and 2–5-km rotation fields are employed to verify both UH and VOR. Before verifying, the MRMS rotation data are interpolated to the 1-km grid using a Cressman interpolation scheme with a 3-km radius of influence.

An object-based verification framework (Skinner et al. 2018) is used to verify the forecasts of mesocyclones for only the Harvey and Springtime cases. While the Michael case had six reported tornadoes during our time of interest, only one tornado occurred within the 1-km domain. For the Summertime case, no tornado occurred during this event and the forecast rotation was weak. Thus, no rotation results are included for these two cases. Like previous studies (Skinner et al. 2018; Lawson et al. 2021), this study uses percentiles to determine the thresholds identifying the strongest rotational areas as objects. Considering the magnitudes of the diagnostic variables vary significantly among different storm types, a fixed percentile value for all cases was not selected for this work. Instead, two percentile thresholds are used for each of the forecast and observed variables, which include 0–2 km averaged vertical vorticity, 0–2-km UH, and 2–5-km UH for the forecasts and 0–2-km and 2–5-km MRMS rotation fields for the observations (Table

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

3). Also, different thresholds (but same percentile) are applied to WoFS-1km and Downscaled-1km, respectively. Sensitivity tests show that the results are consistent when either the Downscaled-1km thresholds or the WoFS-1km thresholds are applied to both runs or to each run respectively. The percentile thresholds are determined using all of the forecasts from all of the members for each case. The 99th and 99.5th percentiles are used for the Harvey case, while 99.9th and 99.95th percentiles are used for the Springtime case. Lower percentiles (99th, and 99.5th) are used for the landfalling TC case because the strongest rotations are located near the eyewall, and a threshold that is too high could miss the rotations in the rainband where tornadoes did occur. The goal is to evaluate the performance of forecasts in predicting strong rotation, rather than to determine uniform thresholds for these variables with such a small sample of cases. Before identifying objects, the MRMS rotation, UH and VOR fields are aggregated to create 30-min rotation tracks. The objects with areas greater than 90 km^2 and that last for at least 15 minutes are retained. Also, adjacent objects with an edge distance of less than 10 km are merged. To match the forecast and the verification objects, a total interest score (TI, Davis et al. 2006a, b; Skinner et al. 2018) is computed:

$$\mathrm{TI} = \left\{ \frac{\left[\frac{\mathrm{cd}_{\max} - \mathrm{cd}}{\mathrm{cd}_{\max}}\right] + \left[\frac{\mathrm{md}_{\max} - \mathrm{md}}{\mathrm{md}_{\max}}\right]}{2} \right\} \left[\frac{t_{\max} - t}{t_{\max}}\right], \qquad (2)$$

in which cd is the centroid distance between the object pair, md is the minimum distance between them, and t is the time difference between them. The maximum allowable threshold for matching is set to 40 km for the centroid ($\mathrm{cd}_{\max}$) and minimum spatial displacement ($\mathrm{md}_{\max}$), and 25 minutes for the time displacement ($t_{\max}$). Matched pairs must have a total interest score greater than 0.2. The matched object pairs are classified as "hits", the unmatched forecast objects as "false alarms", and unmatched verification objects as "misses"; all of which can be used to create the elements of 2 × 2 contingency table. By aggregating hits (a), false alarms (b) and misses (c) from all the members and forecast times, the contingency metrics, probability of detection [POD = $a/(a+c)$], false alarm ratio [FAR=$b/(a+b)$], frequency bias [bias=$(a+b)/(a+c)$], and critical success index [CSI=$a/(a+b+c)$] are computed and are then depicted in a performance diagram (Roebber 2009).

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Table 3 Percentile values of the low- and mid-level rotations from the 1-km forecasts, and MRMS observations

Case	Percentile	Variable	Downscaled -1km	WoFS-1km	Variable	MRMS 1 km
Harvey	99	0–2 km VOR	3.04	3.5	0–2 km rotation	3.0
	99.5	(10^{-3} s^{-1})	4.2	4.7	(10^{-3} s^{-1})	4.1
Springtime	99.9	0–2 km UH	19.56	17.7	0–2 km rotation	2.7
Low-level	99.95	(m^2 s^{-2})	28.28	25.6	(10^{-3} s^{-1})	3.3
Springtime	99.9	2–5 km UH	81.80	74.6	2–5 km rotation	3.3
Midlevel	99.95	(m^2 s^{-2})	131.62	118.1	(10^{-3} s^{-1})	3.9

4. Results and discussion

a. 1-km DA sensitivity experiments

Sensitivity experiments with different combinations of radar observation errors, horizontal covariance localization radii and clear-air reflectivity levels are conducted for the WoFS-1km DA (Table 4). The horizontal covariance localization radius represents the influence radius of each observation, and the localization function is based on the fifth-order correlation function from Gaspari and Cohn (1999). In WoFS, random additive noise is introduced to the *u*, *v*, θ, and *Td* variables with the respective standard deviation of 0.5 m s^{-1}, 0.5 m s^{-1}, 0.5 K and 0.5 K for all the posterior ensemble members where observed reflectivity exceeds 35 dBZ and the difference between the observed and the priors exceed 10 dBZ (Dowell and Wicker 2009; Sobash and Wicker 2015). The horizontal and vertical length scales of the smoothed perturbations are 9 and 3 km, respectively. Sensitivity experiments show that introducing more additive noise during 1-km DA cycles creates too much spurious convection (not shown), and therefore, the additive noise is turned off during 1-km DA. As listed in Table 4, different observation errors are tested in experiments Z7V3_H12 (7 dBZ for reflectivity, 3 m s^{-1} for radial velocity) and Z5V2_H12 (5 dBZ and 2 m s^{-1}, respectively), both with a horizontal localization (HL) radius of 12 km. Experiment Z5V2_H6, which is identical to the experiment Z5V2_H12 except with a smaller HL radius of 6 km, is conducted to test the HL radius for the 1-km DA. To suppress the significant spurious convection associated with the 1-km forecast (not shown), an additional experiment with 8 levels of clear-air reflectivity (Z5V2H12_8ZR), instead of 4 levels as in the baseline WoFS, is conducted.

Figure 3 shows the 1-km DA observation-space diagnostic statistics of radar reflectivity exceeding 15 dBZ and radial velocity for the Summertime case (18 June 2019) during the first 5-

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

h DA-cycling window. The increase of RMSI and bias of reflectivity (Figs. 3a,b) between 1615 UTC to 1715 UTC 18 June 2019 is attributed to the rapid development of storms and a low ensemble-mean reflectivity on average. The comparison between experiments Z7V3_H12 (black curves) and Z5V2_H12 (green curves) shows that using smaller observation errors generates forecasts and analyses with both lower bias and RMSI (Figs. 3a,b). Correspondingly, the total ensemble spread is smaller in Z5V2_H12, partly owing to the lower observation errors. Overall, the prior RMSI and total spread in Z5V2_H12 are comparable in magnitude (Figs. 3b,c,f,g), and the corresponding value of CR approaches 1.0 in the later period of DA time window (Figs. 3d,h), indicating that the ensemble variance is a good approximation of the forecast error variance for the assumed observation errors of reflectivity and radial velocity. The results of experiments Z5V2_H12 (green curves) and Z5V2_H6 (blue curves) show that a 12-km HL radius yields lower bias and RMSI than a 6-km HL radius. Note that the plots of experiments Z5V2H12_8ZR (red curves) and Z5V2_H12 nearly overlap (Fig. 3), which is likely due to only computing observation-space diagnostics within the areas of *observed* reflectivity exceeding 15 dBZ. The clear-air reflectivity observations would have a much larger impact on areas of spurious model storms.

The 3-hr forecast skill of these sensitivity experiments are further evaluated by computing the eFSS of composite reflectivity (Fig. 4). The results show that the experiment Z5V2H12_8ZR performs slightly better than others, though Z7V3_H12 has higher eFSS for composite reflectivity exceeding 25 dBZ in the first 45 mins (Fig. 4a). Both the observation-space diagnostics during DA cycles and the eFSS score during the 3-h forecast indicate that experiment Z5V2H12_8ZR yields the best performance. Sensitivity tests for the other three cases yield similar results (not shown). Thus, the DA configuration of the Z5V2H12_8ZR experiment is used in the WoFS-1km experiment. The different radar DA parameters used in the WoFS-3km and WoFS-1km are summarized in Table 5.

Table 4 1-km radar DA experiments

DA experiments	Reflectivity and radial velocity observation error	Horizontal covariance localization radius	Clear-air reflectivity vertical levels
Z7V3_H12	7 dBZ; 3 m s^{-1}	12 km	1, 3, 5 ,7 km MSL
Z5V2_H12	5 dBZ; 2 m s^{-1}	12 km	1, 3, 5 ,7 km MSL
Z5V2_H6	5 dBZ; 2 m s^{-1}	6 km	1, 3, 5 ,7 km MSL

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Z5V2H12_8ZR	5 dBZ; 2 m s^{-1}	12 km	1, 2, 3, 4, 5, 6, 7, 8 km MSL

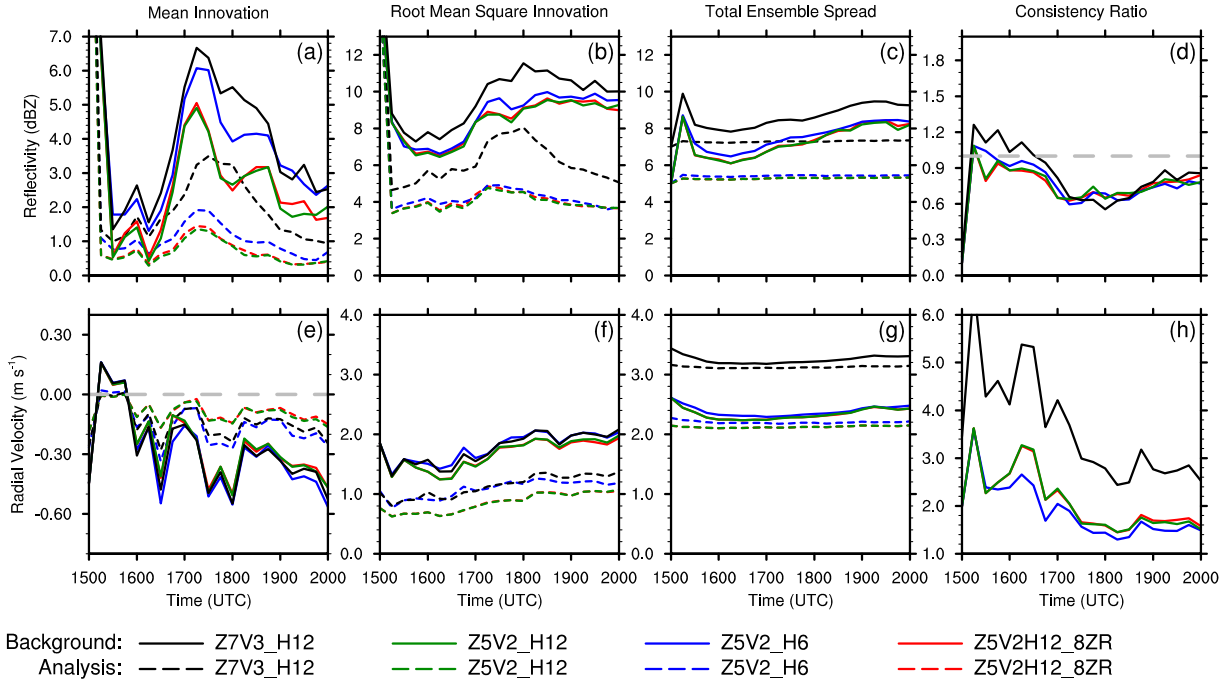
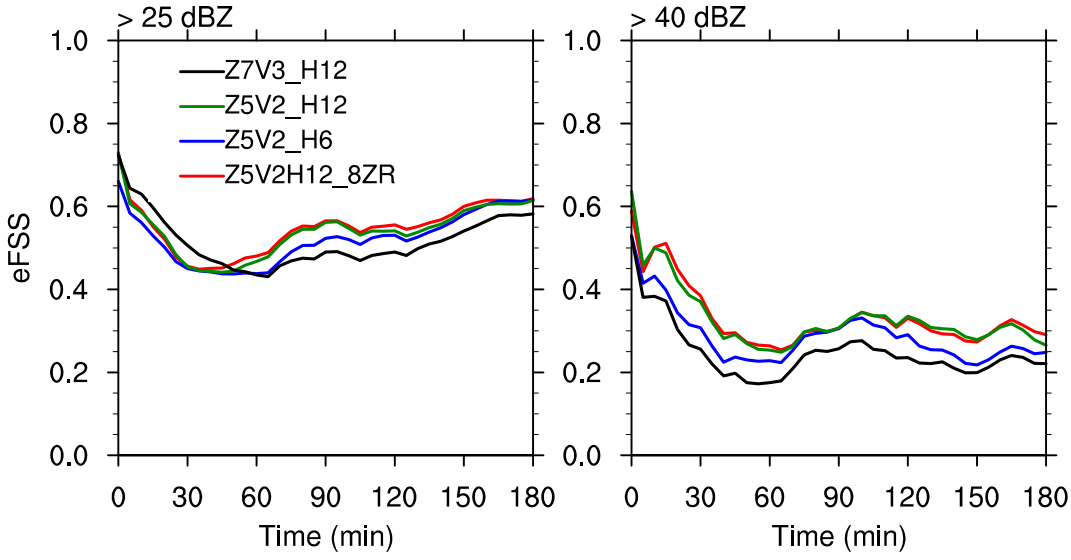


Fig. 3 Observation-space diagnostics, including mean innovation, root-mean-square innovation, total ensemble spread, consistency ratio for background and analysis reflectivity >15 dBZ (a–d) and radial velocity (e–h) for 1-km data DA experiments from 1500 UTC to 2000 UTC 18 June 2019.



Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Fig. 4 Ensemble fractions skill score (eFSS) of composite reflectivity > 25 dBZ and >40 dBZ of the 3-h forecasts initialized at 1900 UTC 18 June 2019 using a neighborhood radius of 12 km for 1-km DA experiments.

Table 5 Radar DA parameters for the WoFS-3km and WoFS-1km

Domain (grid spacing)	Reflectivity resolution/error	Clear-air reflectivity resolution/error/levels	Radial velocity resolution/error	Horizontal localization radius	Reflectivity-based additive noise
D01 (3 km)	5 km; 7 dBZ	15 km; 5 dBZ; 4 levels	5 km; 3 m s^{-1}	18 km	Yes
D02 (1 km)	3 km; 5 dBZ	6 km; 5 dBZ; 8 levels	3 km; 2 m s^{-1}	12 km	No

b. 1-km DA observation-space diagnostics

The reflectivity and radial velocity observation-space diagnostics from the WoFS-1km for all four cases are shown in Figure 5. In general, the large mean innovation and RMSI for both reflectivity and radial velocity sharply decrease after the first couple of DA cycles, indicating that assimilating dense radar data and other observations can quickly reduce the model error. Convection initiation at ~1645 UTC 18 June 2019 and ~1745 UTC 5 May 2020, are underpredicted by the model, causing the increase of the positive mean innovation and RMSI of radar reflectivity at those times (Figs. 5e,g). The RMSI of radial velocity also increases when storms develop and then becomes quite stable (Figs. 5f,h). The total spread curves in Fig. 5 are quite close to the observation error line (e.g., Figs. 5a,b,c,d,f,h), indicating small model ensemble spread. The prior RMSI and total spread are comparable in magnitude for each of the four cases, and the corresponding CRs approach 1.0 in the later period of DA time windows, which are desirable. The overall diagnostic statistics suggests that the 1-km ensemble DA system has reasonable and stable performance for all four cases.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

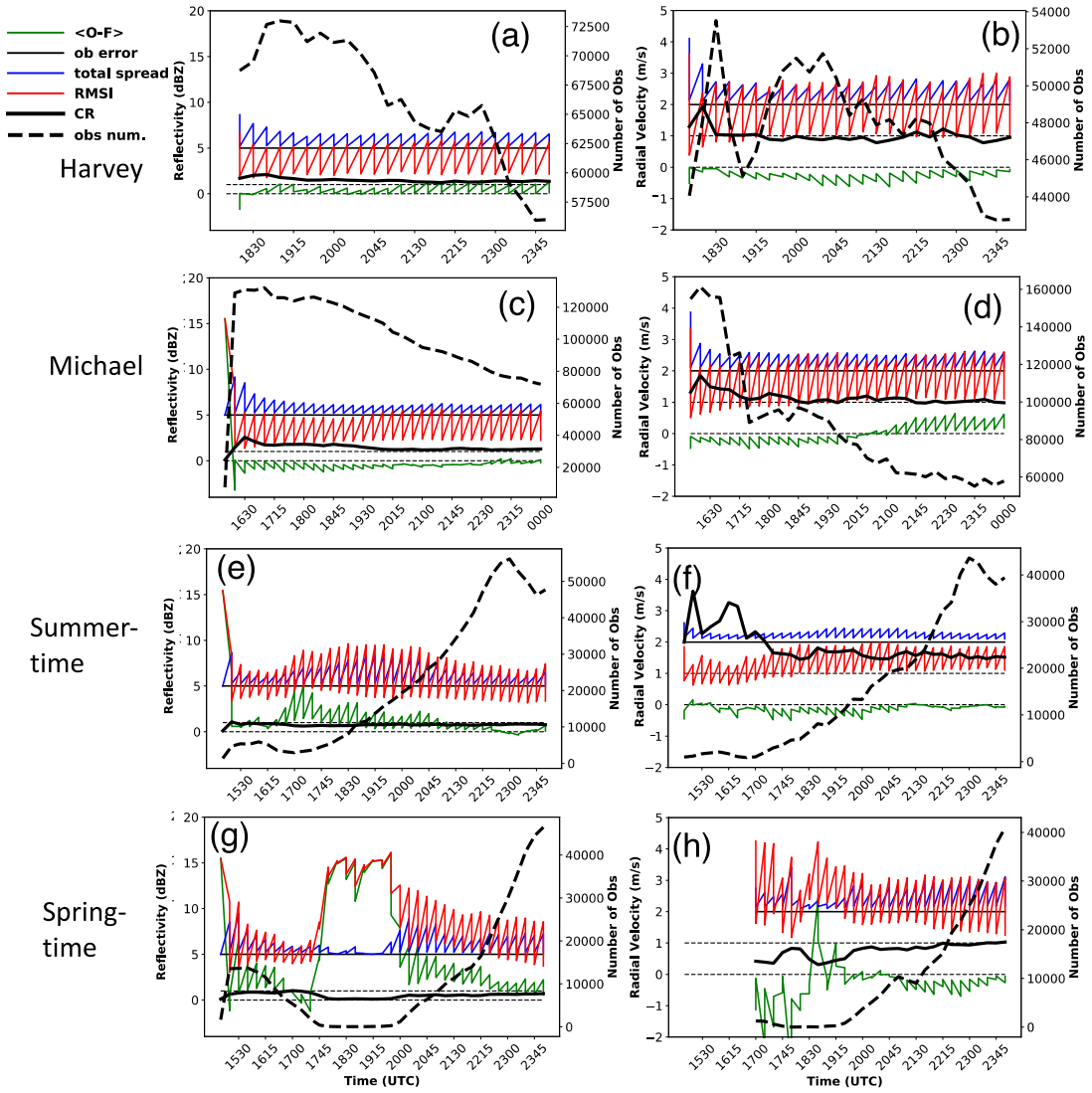


Fig. 5 Mean innovation (<O-F>), root-mean-square innovation (RMSI), observation error (ob error), total spread, consistency ratio (CR) of reflectivity > 15 dBZ (a,c,e,g) and radial velocity (b,d,f,h), and number of assimilated observations (right axis) from the 1-km DA of four cases. Two thin dashed lines denote 0 and 1 values on the left axis. The sawtooth patterns of <O-F>, RMSI and total spread are due to the plotted prior and posterior statistics.

c. Composite reflectivity forecasts

The performance of WoFS-1km and Downscaled-1km experiments are first subjectively evaluated using observations. Figure 6 and Figure 7 show the MRMS composite reflectivity and the probability matched mean (PMM) composite reflectivity from the ensemble analyses and 1-h forecasts, respectively, at their representative times for each case. The probability matching

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

assumes that the most likely spatial representation is given by the ensemble mean and restores the amplitude characteristics of the full ensemble to the ensemble mean (Ebert 2001).

Hurricane *Harvey* reached category 4 status on the Saffir-Simpson scale by 0000 UTC 26 August while it approached the middle Texas coast. It made landfall near Rockport, TX at 0300 UTC 26 August and stalled near the coast for four days. For this case, both experiments accurately depict the hurricane eyewall and rainbands (Figs. 6a,e,i). In the 1-h forecasts, both WoFS-1km and Downscaled-1km overpredict the reflectivity of the convection near the eyewall and in the northeast rainband, but underpredicts the western rainband (Figs. 7a,e,i). Both experiments produce the rainband intersecting the Texas coast to the northeast of the hurricane and the supercells embedded in the northern rainband, but the rainband in WoFS-1km extends further eastward into the Gulf of Mexico like in the observations (Fig. 7i).

Hurricane *Michael* (10 October 2018) moved northward and turned northeastward while approaching the northern Gulf Coast, making landfall near Tyndall Air Force Base in Florida near 1730 UTC on 10 October. By that time, *Michael* reached category 5 hurricane status with maximum sustained winds estimated up to ~140 kt. *Michael* also caused several tornadoes in Florida, Georgia, South Carolina and Virginia during the next several days as it moved farther inland. For Hurricane *Michael*, the eyewall is well depicted in both experiments (Figs. 6b,f,j). The WoFS-1km experiment even reproduces the narrow outer eyewall on the southwest side with adequate representative coverage of strong reflectivity, indicating the advantage of denser radar DA into the finer-resolution system. However, both experiments slightly overpredict the spatial extent of the higher reflectivities especially as forecasts advance in time and the WoFS-1km has the eye located slightly to the east of the observed eye (Figs. 7b,j).

From 18 to 20 June 2019, storms across the eastern United States caused notable rainfall amounts up to seven inches, leading to instances of flash flooding in parts of New Jersey, Pennsylvania, New York, Virginia and areas near Washington D. C.. For the Summertime case, the WoFS-1km analysis is somewhat better than Downscaled-1km due to the former producing less spurious convection in central Virginia (Figs. 6c,g,k). This result is likely because denser clear-air reflectivity is assimilated into the WoFS-1km. Even though both experiments have small areas of spurious convection near the southern edge of the domain and in central Virginia along with the convection displacement errors, Downscaled-1km has larger and more numerous areas of intense convection (Figs. 7c,g,k).

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Through the afternoon hours of 5 May 2020, a supercell moved east-southeastward along the border between South Carolina and North Carolina. The supercell thunderstorm produced four tornadoes, including one EF-2 tornado in Great Falls, South Carolina, and multiple large hail and strong wind events. For this Springtime case, both 1-km experiments generate accurate 2200-UTC analyses (Figs. 6d,h,l). They are also able to predict the intensity and the location of the main line of convection reasonably well, but overpredict the convection in the west portion of the domain behind the line (Figs. 7d,h,l).

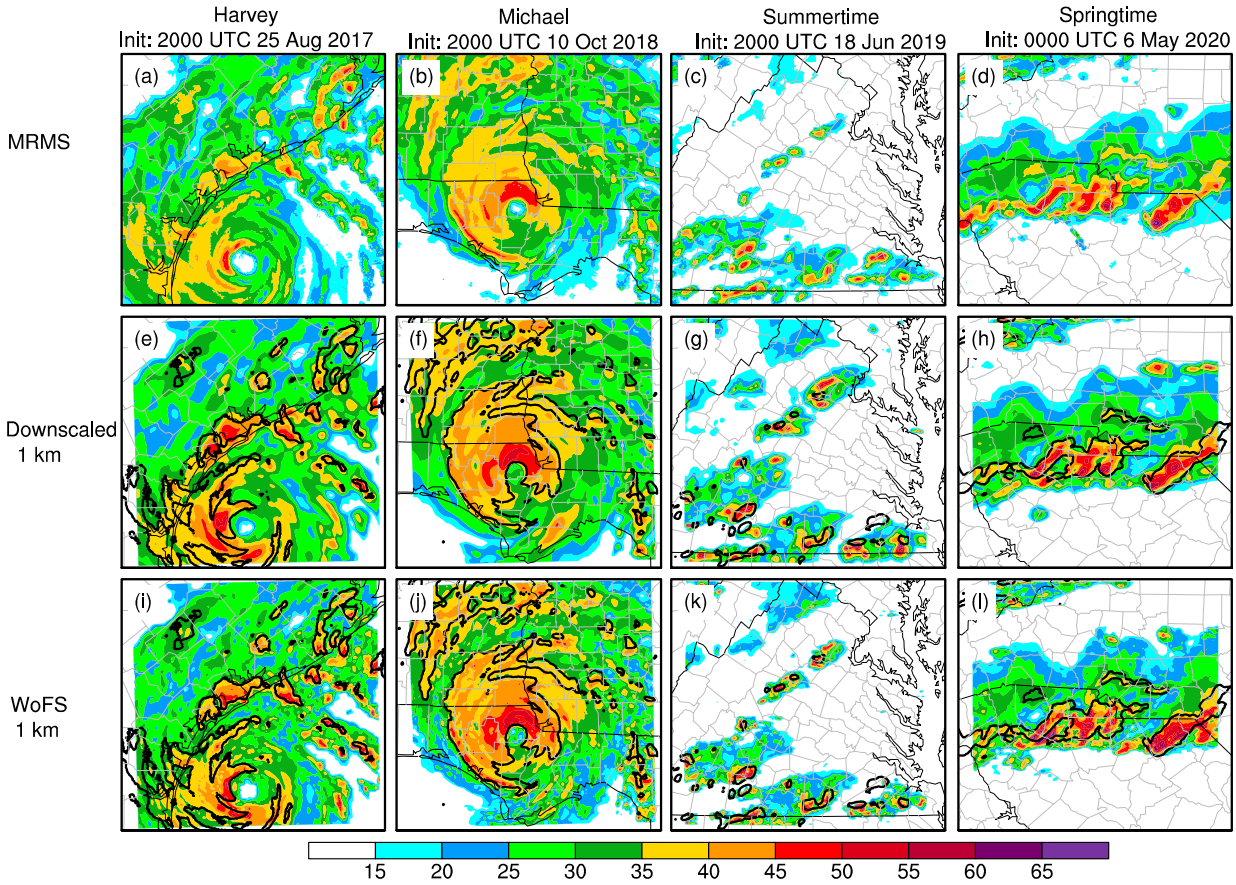


Fig. 6 Observed MRMS composite reflectivity (units: dBZ; a–d,) and simulated probability matched mean composite reflectivity from 3-km (e–h) and 1-km (i–l) analyses at 2000 UTC 25 Aug 2017, 2000 UTC 10 Oct 2018, 2000 UTC 18 Jun 2019, and 0000 UTC 6 May 2020. Black contours represent the 35 dBZ value from MRMS.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

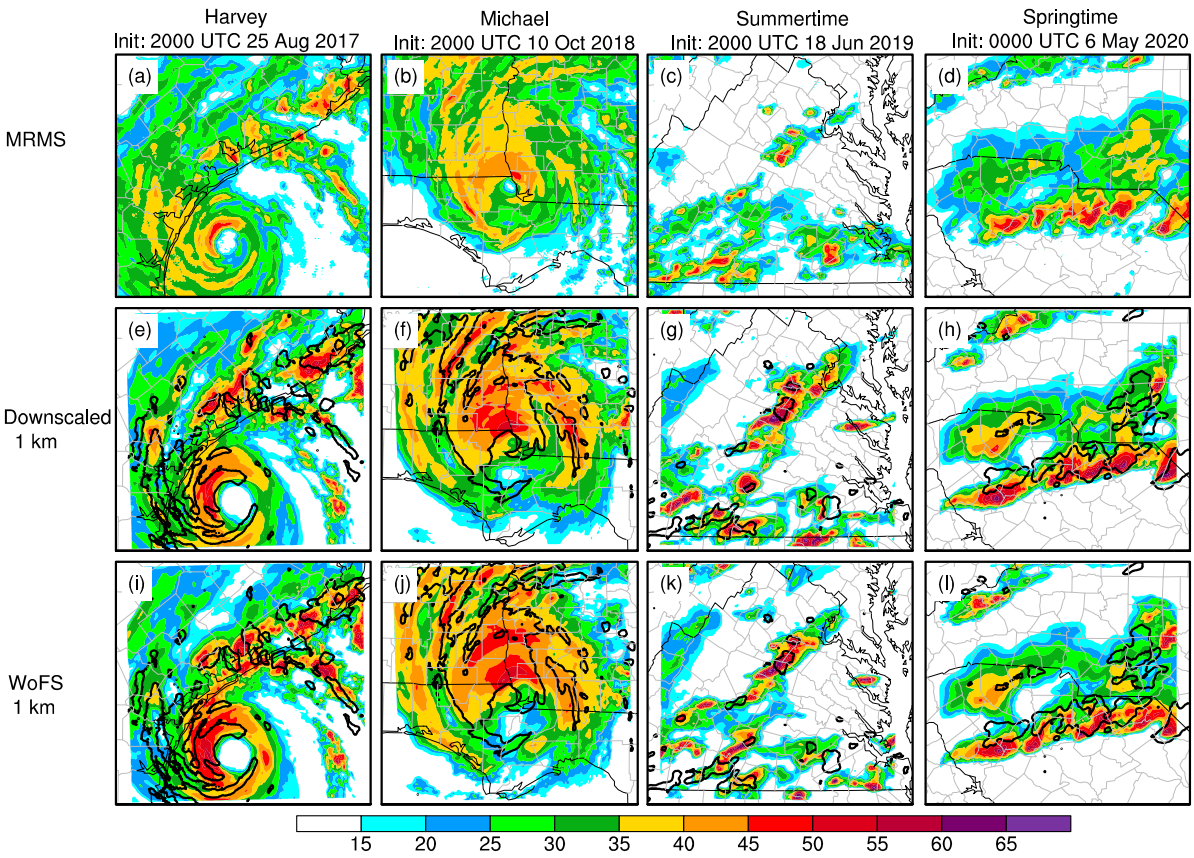


Fig.7 Similar to Fig. 6, except from the 1-h forecasts initialized at 2000 UTC 25 Aug 2017, 2000 UTC 10 Oct 2018, 2000 UTC 18 Jun 2019, and 0000 UTC 6 May 2020.

Objective evaluation of composite reflectivity is conducted by computing eFSS for the four cases (Fig. 8). To eliminate the possible impact of model bias, the eFSS of composite reflectivity exceeding the 75th, 95th, and 99th percentile thresholds are computed using forecast history files every 5 min from all the initializations and all ensemble members (Table 6). The time series of eFSS shows that WoFS-1km generally outperforms Downscaled-1km at forecasting higher reflectivities (99th percentiles) in the two TC events, while it is less skillful for weaker reflectivities (75th percentile). The impact of 1-km DA is quite neutral for the 95th percentile reflectivity in the *Michael* case. For the Summertime event, WoFS-1km shows marginal improvement, with the most sustained improvement through the 3-hr forecast for the 99th percentile. For the Springtime case, even though WoFS-1km generates higher eFSS at the analysis time, its forecast is worse for all thresholds (Figs. 8h,l), probably due to less intense convective cores and larger storm displacement

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

errors within the leading line of convection (Figs. 7h,l). The results indicate that WoFS-1km's performance varies over different cases, and is likely more beneficial for high reflectivity forecasts.

Table 6 Percentile thresholds of composite reflectivity and hourly accumulated precipitation for forecast experiments and MRMS observations

Experiment	Composite reflectivity (dBZ)			Hourly accumulated precipitation (mm h^{-1})				
	75th	95th	99th	75th	90th	95th	99th	99.5th
WoFS-1km	26.3	41.7	49	1.8	7.1	11.2	22.6	28.2
Downscaled-1km	26.9	41.7	49	2.03	7.2	11.2	21.9	27.2
MRMS	25.3	36.7	42	2.3	6.9	11.3	23	29

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

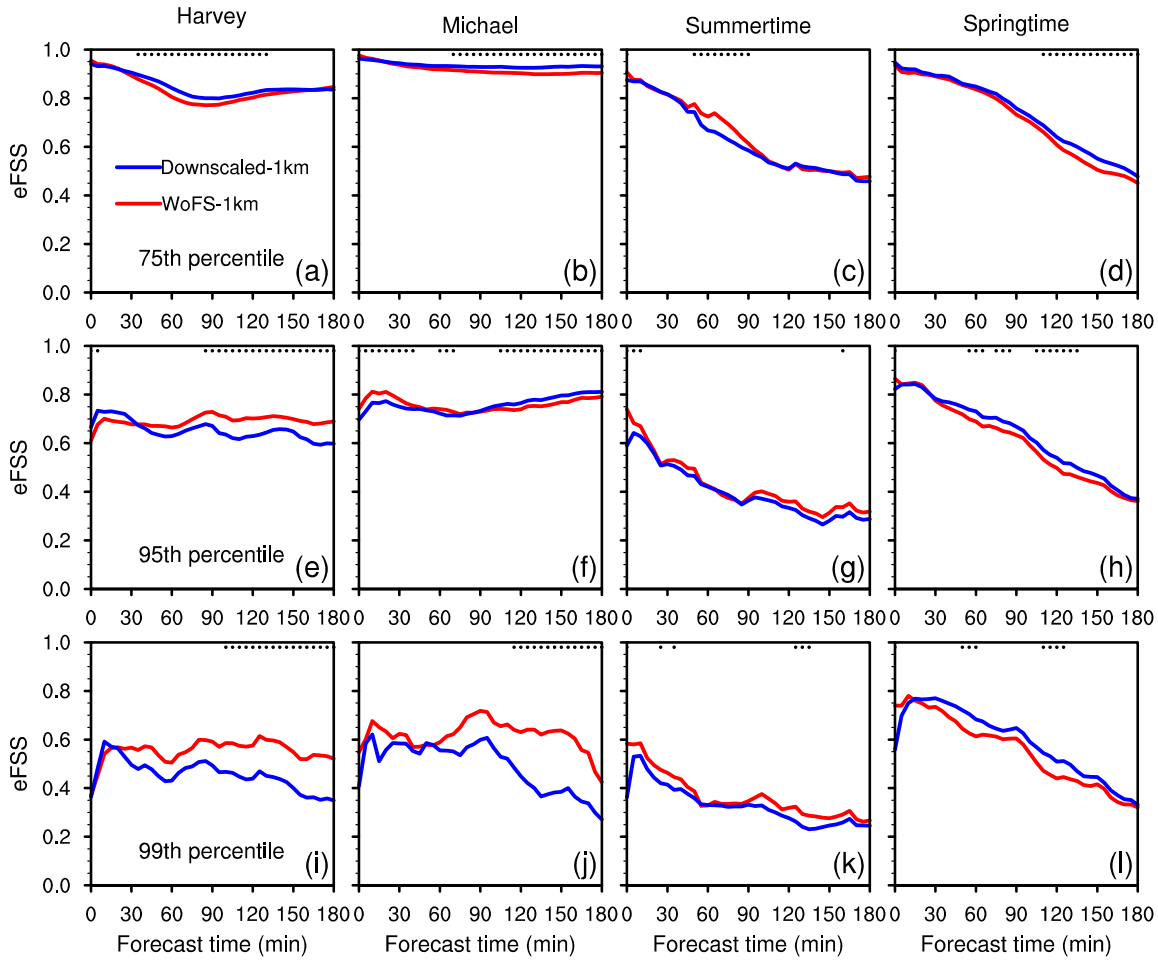


Fig. 8 Ensemble fractions skill score (eFSS) of forecasted composite reflectivity exceeding different percentile thresholds from WoFS-1km and Downscaled-1km using a neighborhood radius of 12 km. The eFSS are aggregated over all the initialization times. The black dots from the top indicate the eFSS differences between the two experiments are statistically significant.

d. Precipitation forecasts

To evaluate the performance of the two 1-km forecast experiments in predicting heavy precipitation, the forecasts of PMM 3-h accumulated precipitation are compared against the MRMS QPE. Hurricane *Harvey* stalled with its center near the Texas coast for four days after landfalling, dropping historic rainfall amounts of more than 60 inches over parts of southern Texas and causing catastrophic flooding. For this case, the extremely heavy precipitation from the forecast initialized at 0000 UTC 26 August is better depicted by the WoFS-1km than the Downscaled-1km when compared to the observations in the same time frame. Specifically, the eyewall precipitation center over the Texas coast and the precipitation center in the northeastern rainbands are slightly more intense and similar to the observations in the WoFS-1km experiment

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

(Figs. 9a,e,i). However, both WoFS-1km and Downscaled-1km vastly underestimate the precipitation in the northeastern rainbands.

Hurricane *Michael* caused widespread rainfall accumulations of 3 to 6 inches and localized rainfall totals in excess of 10 inches after its landfall, resulting in some freshwater flooding as it moved across the southeastern United States. Similar to the *Harvey* case, WoFS-1km predicts a more intense eyewall precipitation center than the Downscaled-1km, with the intensity more consistent with the observations (Figs. 9b,f,j). However, both experiments fail to represent the intense precipitation structure around the northwestern/western portions of the eyewall and the narrower precipitation bands wrapping around the southern extent of the eye. Away from the eyewall, both experiments slightly underpredict the northern outer rainband.

For the Summertime case, the observed precipitation center near the southern edge of Virginia caused a couple of flash flooding events at ~2230 UTC on 18 June, 2019 according to the NWS Weather Prediction Center flash flood reports with 3-hourly precipitation totals up to 50 mm. The 3-h forecast PMM precipitation from both experiments shows an overestimation of rainfall in the northern part of the domain and an underestimation of rainfall in the southwestern portion of the domain compared to the observations (Figs.9 c,g,k). WoFS-1km offers some marginal improvement over Downscaled-1km since Downscaled-1km falsely generates more areas of spurious precipitation centers, especially in the southeastern portion of the domain. For the Springtime case, both experiments underpredict the precipitation amount in the western portion of the domain and expand the precipitation too far to the east compared to the observations (Figs. 9d,h,l). Downscaled-1km overall has a more continuous strong precipitation area like the observations than the WoFS-1km.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

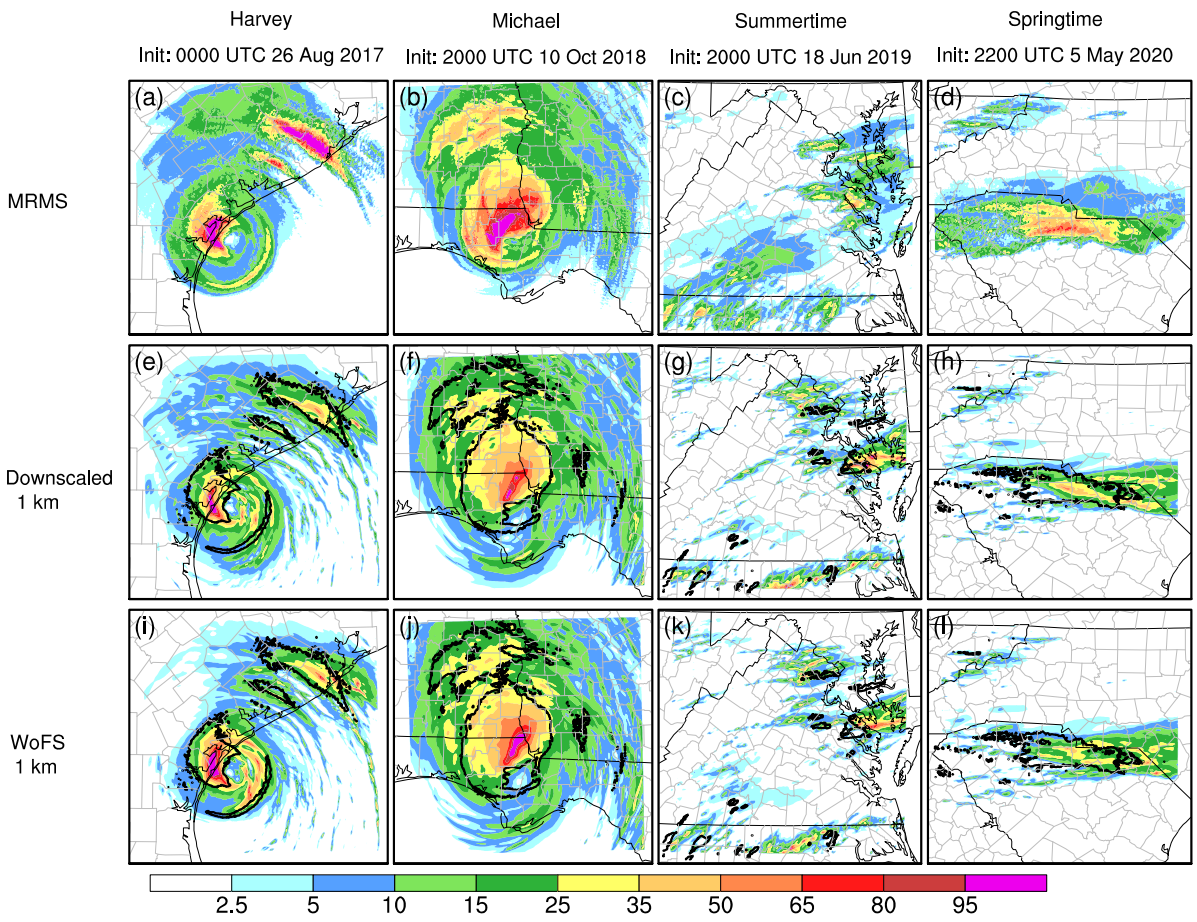


Fig. 9 The 3-h accumulations of MRMS QPE (a–d) and the probability matched mean (PMM) 3-h accumulated precipitation (units: mm) from downscaled-1km (e–h) and WoFS-1km (i–l) forecasts initialized at 0000 UTC 26 Aug 2017, 2000 UTC 10 Oct 2018, 2000 UTC 18 Jun 2019, and 2200 UTC 5 May 2020. Black contours represent the observed 25 mm values.

The two 1-km experiments produce very similar precipitation coverage fractions for smaller thresholds of 0.25–10 mm h^{-1} (Fig. 10). The deviation between the two experiments increases with higher thresholds. Specifically, the two experiments forecast the precipitation coverage well at lower thresholds for the two landfalling TC cases (Figs. 10a,b). While the forecasts from both experiments underpredict at higher thresholds, WoFS-1km creates slightly larger areas of heavier precipitation, which is more consistent with the observations in the terms of areal coverage, than the Downscaled-1km experiment. Both experiments slightly underpredict at lower thresholds for the Spring and Summertime cases, while overpredict at higher thresholds. However, the forecast of areal coverage at higher rainfall thresholds of WoFS-1km is closer to the observed areal coverage (Figs. 10c,d). The aggregated result shows that while both experiments

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

underestimate precipitation greater than 20 mm h^{-1}, WoFS-1km performs slightly better than Downscaled-1km in representing the coverage of higher rainfall (Fig. 10e).

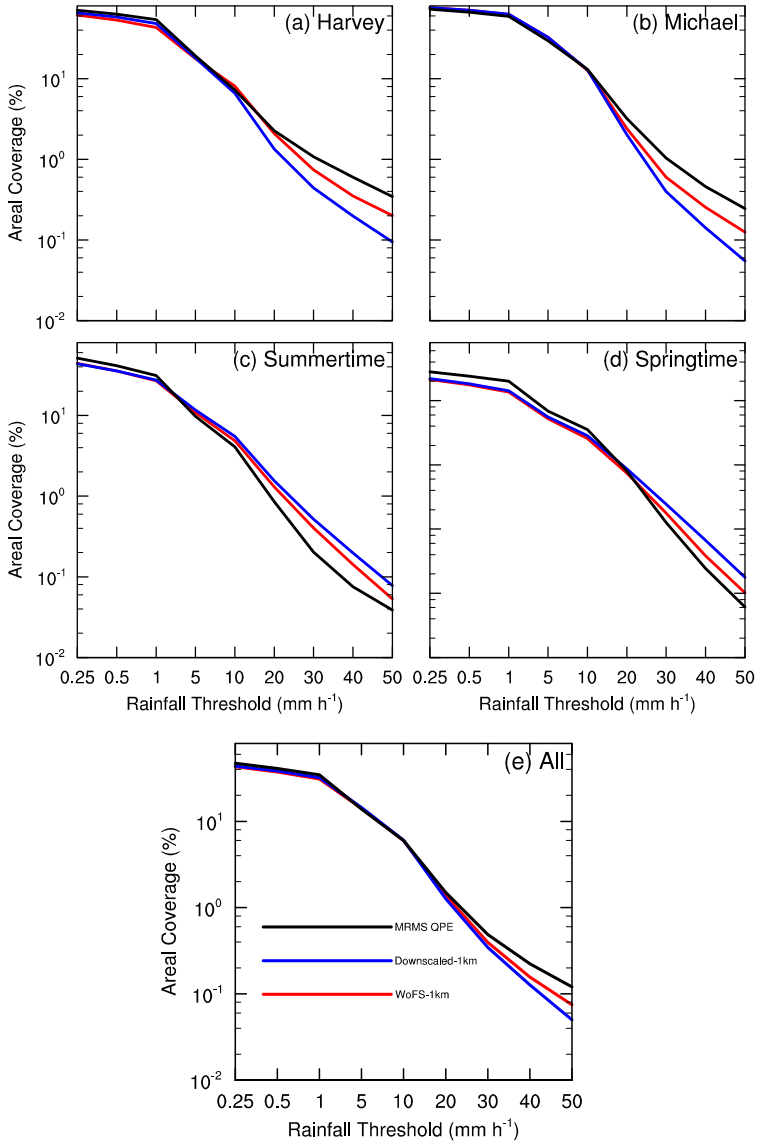


Fig. 10 Fractional areal coverage (%) of 1-h accumulated precipitation exceeding different thresholds from (a–d) each case and (e) all four cases. Hourly accumulated precipitation is aggregated over all initializations and forecast hours.

To complement the above subjective verification of the precipitation forecasts from both 1-km experiments, the objective verification metric eFSS is computed to evaluate the spatial placement accuracy. The eFSS is computed for the 1-h accumulated precipitation forecast aggregated over all of the initialization and forecast times with different percentile thresholds

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

(Table 6) and neighborhood radii for each case. Percentile rather than accumulation thresholds are used to reduce the impact of precipitation bias in the calculation of eFSS of the two 1-km experiments (as shown in Fig. 10). As shown in Fig. 11, the performance of the two experiments is case sensitive. Specifically, for Hurricane *Harvey*, the WoFS-1km is more skillful than Downscaled-1km in predicting heavier rainfall exceeding the 90th, 95th, 99th and 99.5th percentiles, but less skillful in predicting moderate rainfall exceeding the 75th percentile (Figs. 11a1−a5). For Hurricane *Michael*, Downscaled-1km outperforms WoFS-1km at the 75th percentile while the difference decreases at higher percentile thresholds (Figs. 11b1−b5). For the Summertime case, the WoFS-1km shows higher eFSS than Downscaled-1km at moderate precipitation (75th percentile) but essentially indistinguishable difference for higher percentiles thresholds (Figs. 11c1−c5). For the Springtime case, Downscaled-1km performs better than WoFS-1km at all percentile values (Figs. 11d1−d5). For all cases, the eFSS difference between the two experiments are not statistically significant at the 95% level probably owing to a small sample size.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

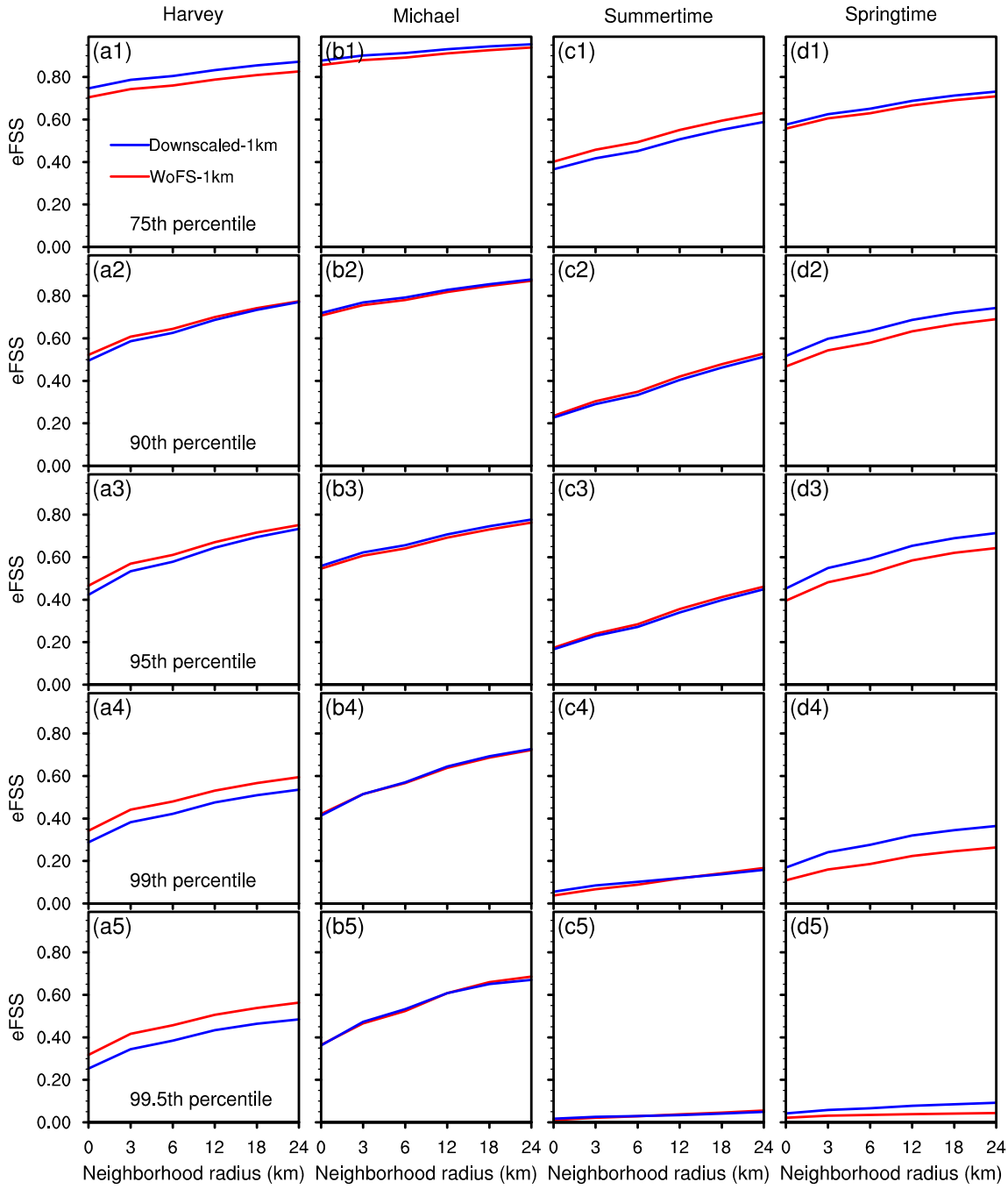


Fig. 11 Ensemble fractions skill score (eFSS) for each neighborhood radius (km) using 1-hourly accumulated precipitation greater than different thresholds from WoFS-1km and Downscaled-1km forecasts. The eFSS are computed using hourly precipitation from all the members, initializations and forecast hours. The eFSS difference between the two experiments are not statistically significant at the 95% level.

The eFSS differences between the two experiments for the first, second, and third-hour precipitation are aggregated together over all four cases and are shown in Fig. 12. WoFS-1km outperforms the Downscaled-1km in predicting heavier precipitation exceeding the 99th and 99.5th

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

percentiles and underperforms for moderate precipitation. The difference between the two experiments gradually increases with longer forecast lead times, indicating that the impact of 1-km DA can last for hours. In fact, during the third forecast hour, WoFS-1km significantly outperforms Downscaled-1km at the higher percentiles while only having slightly less skillful forecasts than Downscaled-1km at the smaller percentiles (Fig. 12c). The significant outperformance of WoFS-1km is primarily due to its increasing outperformance in *Harvey* and *Michael*, and its slightly reducing underperformance in the Springtime case with longer lead time at higher percentiles. The increasing difference with forecast lead time between the two experiments is likely due to the difference of reflectivity forecast skills in two heavy precipitation-producing TCs as shown in Figs. 8i,j. It is probably because TC rainband and eyewall structures are better represented in WoFS-1km at early forecast times (e.g., Figs. 6,7). In reality, TC convection continued to develop during the forecast periods. With better-resolved TC structures, convective cells in WoFS-1km forecasts are developed more realistically than those in Downscaled-1km and thus, the difference in skill between the two experiments increases as forecast lead time increases.

Overall, WoFS-1km produces heavy precipitation with a smaller bias in area coverage than Downscaled-1km. Nevertheless, the benefit of WoFS-1km in reducing precipitation displacement also depends on individual cases, lead times, and rainfall intensities. The results of precipitation and reflectivity forecasts are generally consistent, except for the *Michael* case, in which WoFS-1km's forecast skill for strong reflectivity is better than Downscaled-1km's forecast skill (Fig. 8j) even though their forecast skills for heavy precipitation are similar (Figs. 11b4,b5).

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

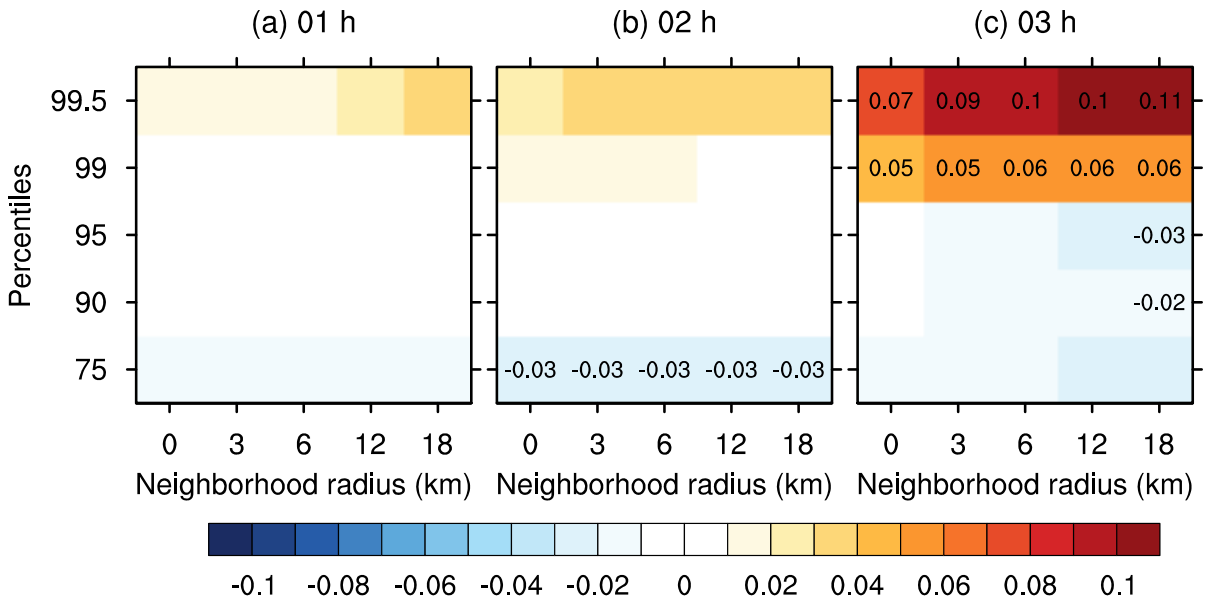


Fig. 12 Ensemble fractions skill score (eFSS) difference (WoFS-1km minus Downscaled-1km) for the first (a), second (b), and third (c) forecast hour accumulated precipitation aggregated over all the initializations greater than different thresholds for each neighborhood radius (km). The overlaying numbers indicate the eFSS differences between the two experiments are statistically significant.

e. Low- and midlevel rotation forecast

Between 1900 UTC 25 Aug and 0300 UTC 26 Aug 2017, Hurricane *Harvey* produced a total of twelve tornadoes. To assess the potential of the WoFS-1km in forecasting low-level rotation in mini supercells within the hurricane rainbands, the 1-h forecast probability swaths of the model-simulated 0–2-km vorticity overlaid by MRMS 0–2-km rotation tracks and tornado reports are shown in Fig. 13. To obtain the probability swath, the maximum vorticity within the 0–1-h period is computed for each member first, and then probabilities of vorticity exceeding certain thresholds (i.e., event occurrence) within a 3-km-radius neighborhood of each grid is computed. The probability here is the neighborhood maximum ensemble probability (NMEP; Schwartz and Sobash 2017). Overall, the WoFS-1km experiment generates higher probability swaths along the coast where several tornadoes occurred compared to Downscaled-1km. In the 1-h forecasts initialized at 2000 UTC and 2300 UTC (Figs. 13), the Downscaled-1km forecasts indicate low probability (≤ 40%) of strong low-level rotation occurrence in the northern rainband where strong rotation swaths and tornadoes were observed, while the WoFS-1km experiment has higher probabilities (> 70%). However, the WoFS-1km experiment still exhibits large

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

displacement errors and noise when compared with the observations. Besides, even WoFS-1km fails to develop any meaningful low-level rotation in the vicinity of the northeastern-most tornado report and the most elongated and contiguous area of MRMS rotation because most ensemble members could not develop strong enough storms there. The notable higher probabilities in WoFS-1km exist throughout the entire 3-h forecast (not shown), indicating that it is not simply a result of faster spin-up when initialized from 1-km analyses (WoFS-1km) than from 3-km analyses (Downscaled-1km).

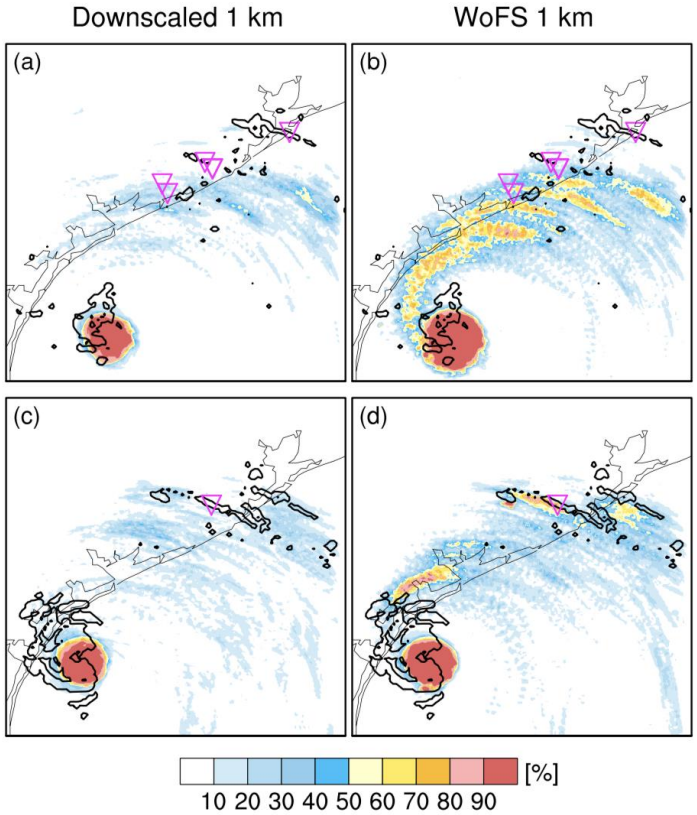


Fig. 13 Neighborhood maximum ensemble probability of simulated 0–2 km vertically-averaged relative vorticity (s^{-1}) exceeding 0.004 s^{-1} within a 3-km-radius neighborhood. The relative vorticity is aggregated within the 0–1-h forecast period beginning at (a,b) 2000 UTC and (c,d) 2300 UTC 25 Aug 2017. The thick black contours are the corresponding low-level MRMS rotation of 0.003 s^{-1} and magenta triangles represent tornado reports during the 1-h forecast period.

There were four tornadoes and several large hail events reported during the entire forecast window for the Springtime case. As shown in Figs. 14, WoFS-1km generates higher probability swaths of strong low- and midlevel updraft helicity than Downscaled-1km over the observed

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

rotation tracks and tornado reports in the eastern part of the domain, although they both miss most of the strong rotational areas to the west. Meanwhile, both experiments produce spurious rotation in the eastern part of the domain, especially in the WoFS-1km experiment. The WoFS-1km tends to generate higher probabilities farther to the south than the Downscaled-1km. Despite the stronger UH in WoFS-1km (Fig. 14), WoFS-1km could not develop a strong enough convective line in the later-initialized forecasts, eventually leading to weaker rotations than that in Downscaled-1km (see Fig. 16 below).

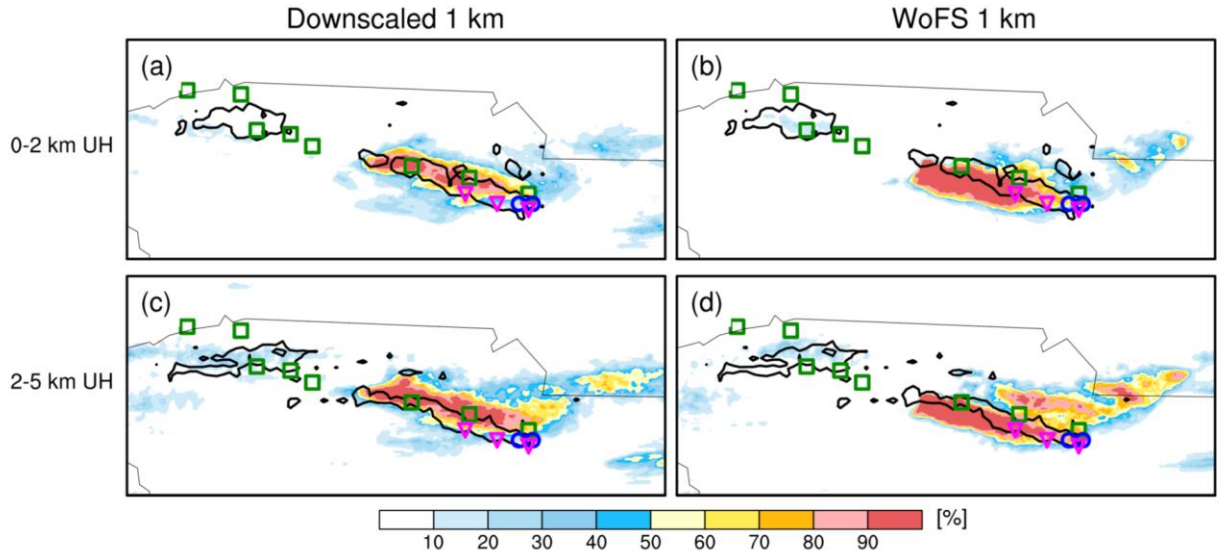


Fig. 14 Similar to Fig. 13, except for (a,b) 0–2 and (c,d) 2–5 km updraft helicity (UH, units: $m^2 s^{-2}$) exceeding 40 $m^2 s^{-2}$ and 130 $m^2 s^{-2}$ respectively and the corresponding low-level and midlevel MRMS rotations of 0.0033 s^{-1} and 0.0039 s^{-1} respectively within the 0–1-h forecast period beginning at 2200 UTC 5 May 2020. The magenta triangles and green squares represent tornado and hail reports respectively during the 1-h forecast period.

The contingency-table elements are calculated using the procedure described in Section 3 and aggregated over all ensemble members, forecast times, and initializations. The overall object-based verification contingency metrics are displayed in Fig. 15. For the *Harvey* case, the two experiments are nearly unbiased in predicting low-level rotations for the 99th percentile threshold while WoFS-1km outperforms Downscaled-1km with a higher CSI. The higher CSI value of WoFS-1km for the 99th percentile threshold is mainly attributed to the lower false alarm ratio, which is partly because Downscaled-1km falsely produces too many large vorticity objects over

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

the ocean (not shown). For the 99.5th percentile threshold, although WoFS-1km has a notable underprediction bias, the two experiments exhibit similar CSI (Figs. 15a).

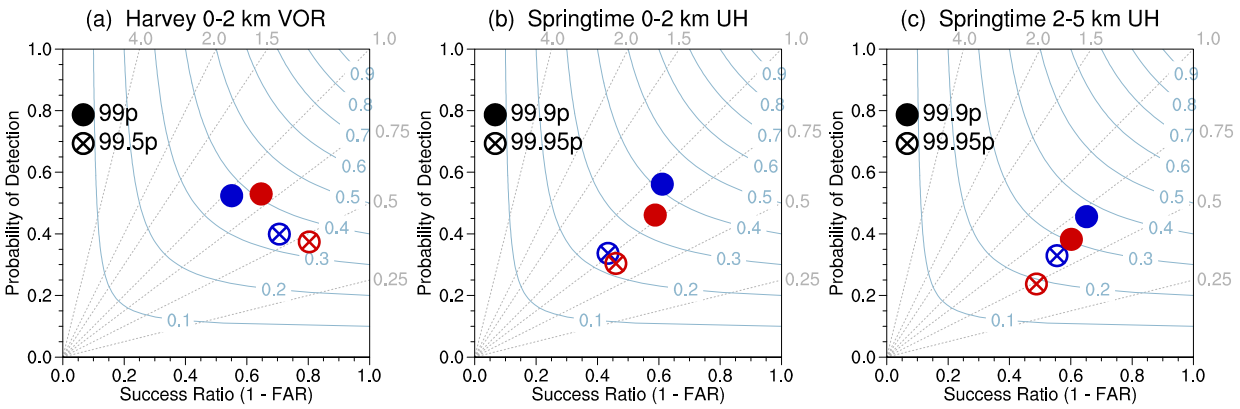


Fig. 15 Performance diagram for (a) 0–2 km VOR objects for *Harvey*, and (b) 0–2 km and (c) 2–5 km UH objects for the Springtime case from WoFS-1km (red marks) and Downscaled-1km (blue marks). The gray dashed lines represent bias scores with labels on the outward extension of the line. The solid curves represent CSI. Score metrics in the performance diagrams are computed by aggregating the table components over all the members, initializations and forecast times. The objects are identified by different percentile values in Table 3.

For the Springtime case, the overall Downscaled-1km forecasts for low- and mid-level rotations are better than those of WoFS-1km (Figs. 15b,c), with higher POD, lower false alarm ratio, and higher CSI, except that the two experiments perform similarly for stronger low-level rotations at a higher percentile (99.95th) threshold (Fig. 15b). The lower forecast skill of WoFS-1km is primarily attributed to it having fewer UH objects in total (Fig. 16a–d). For the 99.9th percentile threshold, a total of 4383 and 5313 0–2-km UH objects are identified over all the initializations, forecast times, and ensemble members in WoFS-1km and Downscaled-1km, respectively. Although the storm morphology appears similar in the 3-km and 1-km analyses, small differences in the heterogeneous low-level environment eventually led to different storm evolutions. Specifically, the analyses of WoFS-1km have stronger cold pools associated with a supercell on the eastern portion of the domain than Downscaled-1km (not shown, similar to Fig. 16e,f). However, ongoing storms on the western portion of the domain are embedded in a warmer analyzed environment in WoFS-1km than in Downscaled-1km. As the forecast advances, the convective cores in WoFS-1km could not develop as vigorously as in Downscaled-1km, and the convective line in WoFS-1km became narrower and discontinuous (see magenta contours in Figs.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

16 e–h), resulting in fewer strong UH objects (Figs. 16b,d), especially in the last two initializations. The reason for the different low-level environments in the analyses from the 1-km and 3-km systems is uncertain, but it is likely to be, at least in part, due to different DA strategies of radar observations. This result highlights the necessity to further tune the DA strategies in WoFS-1km.

The case-dependent performance of WoFS-1km and Downscaled-1km warrants further exploration. Since the radar coverage is sparser near the coast than farther inland, one possible reason is that assimilating denser radar observations is more advantageous for cases located in radar-sparse areas (i.e., landfalling hurricanes near the coast) than cases located inland. For example, for cases in radar-sparse areas, assimilating higher-resolution radar observations as in the WoFS-1km possibly improve the representation of storm structures in the analyses. Since the evolution of TCs are mainly dominated by large-scale processes, they are more likely to benefit from the better analyzed storm structures. In comparison, convective storms like the Springtime case are sensitive to finer-scale, near-storm thermodynamic and dynamic environments, which makes the situation more complicated.

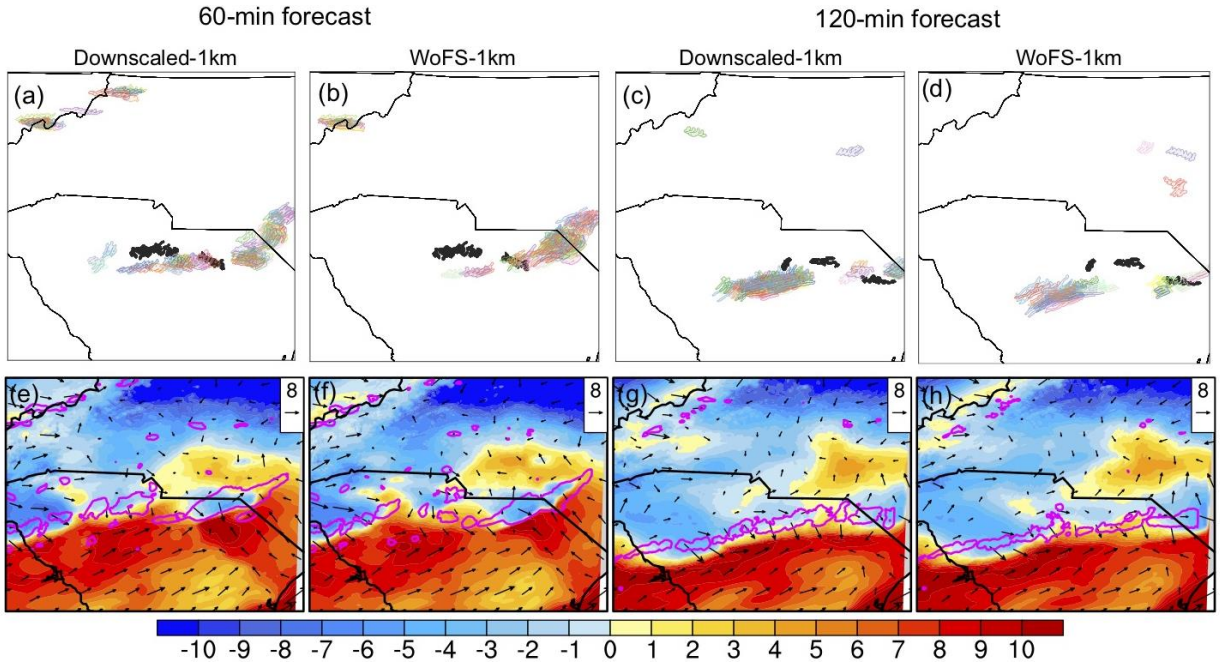


Fig. 16 (a–d) 0–2 km 99.9th percentile UH objects (colored) from each ensemble member and (e–h) the perturbation of the ensemble-mean equivalent potential temperature at the lowest 100-hPa (colored, units: °C) and 10-m wind vectors (units: m s^{-1}) from the (a,b,e,f) 60-min and (c,d,g,h) 120-min forecasts initialized at 2300 UTC 5 May 2020. The MRMS low-level rotation objects in (a–d) are black shaded. The magenta contours in (e–h) represent the 40-dBZ probability matched mean composite reflectivity.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

5. Summary and Conclusions

In this study, an experimental 1-km Warn-on-Forecast DA and forecast system (WoFS-1km) is developed and tested using four high-impact hazardous weather events with different storm types. Two 1-km ensemble forecast experiments are conducted. In one experiment, forecasts are initialized from WoFS-1km analyses. In another experiment, forecasts are initialized from the downscaled WoFS-3km, referred to as Downscaled-1km. The ensemble forecasts of composite reflectivity, precipitation and rotation (i.e., updraft helicity and relative vorticity) from WoFS-1km and Downscaled-1km are assessed using observations and are compared to evaluate the impact of DA at 1-km Δx.

The comparison between WoFS-1km and Downscaled-1km suggests that DA at 1-km Δx benefits the short-term forecast of two cases – one landfalling TC case and the Summertime thunderstorm event. The impact of 1-km DA is neutral on the other TC case, but results in degraded forecast skill in the Spring thunderstorm event. Subjective evaluation shows that WoFS-1km produces better reflectivity forecasts with more complete structure details in one landfalling TC and less spurious convection in the Summertime event compared to that from Downscaled-1km. In general, WoFS-1km shows higher forecast skill in predicting high reflectivity in three cases. Meanwhile, the WoFS-1km experiment performs better at predicting these selected flash flooding events with a smaller bias for heavy rainfall. However, WoFS-1km only marginally improves the spatial placement of heavy precipitation and slightly degrades the skill for weak precipitation in these flash flooding cases. The impact of 1-km DA on precipitation forecasts becomes more apparent with longer forecast lead times. The WoFS-1km improves the forecast skill of strong low-level rotation in the Hurricane *Harvey* case with a lower false alarm ratio. However, the forecast skill for the Springtime event is degraded with 1-km DA likely because of fewer strong convective cells and fewer rotation objects being generated.

Overall, results in this paper only suggest conditional benefits from the WoFS-1km. However, with just four cases being used to test both systems, the results may be impacted by the unique characteristics of each of those cases. A larger sample of cases is needed to further evaluate and generalize the overall performance of the finer-resolution WoFS-1km system. Although Potvin et al. (2017) found that the 2-h convective forecast is relatively insensitive to the Δx of ICs

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

for a few cases, the mixed impact of using higher-resolution ICs in this study is not fully consistent with their findings. Their findings are based on the assumptions that the ICs are perfect and the only source of forecast error is the IC resolution error. However, the newly-developed WoFS-1km in this study is different from the WoFS-3km not only in Δx, but also in observation and DA strategies. These differences between the DA systems ultimately have an impact on the forecasts and their errors. Moreover, this experimental WoFS-1km is in its infancy, and research efforts are currently underway to tune and test the system using a wide variety of hazardous weather events across the continental United States.

Acknowledgements. Funding for this research was provided by NOAA/NSSL FY 2018 Director's Discretionary Research Fund and NOAA-University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. Local computer assistance was provided by Gerry Creager and Joshua Laferriere. We thank Anthony Reinhart and Thomas Jones for processing radar observations. We thank Patrick Skinner for providing the object-based verification scripts and Thomas Jones for kindly providing informal review of this manuscript.

Data availability statement

All experiment datasets used for this study are stored on the NSSL high-performance computing server and the data can be made available upon request.

Reference:

Adlerman, E. J. and Droegemeier, K. K., 2002: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters. *Mon. Wea. Rev.*, **130**, 2671–2691, https://doi.org/10.1175/1520-0493(2002)130<2671:TSONSC>2.0.CO;2.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus, **61A**, 72–83, https://doi.org/10.1111/j.1600-0870.2008.00361.x.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Britt, K. C., P. S Skinner, P. L. Heinselman, K. H. Knopfmeier, 2020: Effects of horizontal grid spacing and inflow environment on forecasts of cyclic mesocyclogenesis in NSSL's Warn-on-Forecast System (WoFS), *Weather and Forecasting*, **35**, 2423-2444, https://doi.org/10.1175/WAF-D-20-0094.1.

Bryan, George H., J. C. Wyngaard, and J. M. Fritsch. (2003). Resolution Requirements for the Simulation of Deep Moist Convection. *Mon. Wea. Rev.*, **131**, no. 10 (October): 2394–2416. https://doi.org/10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2.

Clark, A. J., and coauthors, 2020: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather, *Bull. Amer. Meteor. Soc.*, **101**, E2022-E2024, https://doi.org/10.1175/BAMS-D-19-0298.1

Davis, C. A., B. G. Brown, and R. G. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

——, ——, and R. G. Bullock, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, https://doi.org/10.1175/MWR3146.1.

Dowell, D., L. J. Wicker, and D. J. Stensrud, 2004: High resolution analyses of the 8 May 2003 Oklahoma City storm. Part II: EnKF data assimilation and forecast experiments. Preprints, 22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., 12.5. [Available on line at http://ams.confex.com/ams/pdfpapers/81393.pdf.]

——, and ——, 2009: Additive noise for storm-scale ensemble data assimilation. J. *Atmos. Oceanic Technol.*, **26**, 911–927, https://doi.org/10.1175/2008JTECHA1156.1.

——, ——, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294, https://doi.org/10.1175/2010MWR3438.1.

——, and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. 28th Conf. on Severe Local Storms, Portland, OR,

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

Duc, L., K. Saito, and H. Seko, 2013: Spatial–temporal fractions verification for high-resolution ensemble forecasts. Tellus, **65A**, 18171, https://doi.org/10.3402/tellusa.v65i0.18171.

Ebert, E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, https://doi.org/10.1175/1520-0493(2001)129,2461:AOAPMS.2.0.CO;2.

Gallo, B. T., and coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment, *Weather and Forecasting*, **32**, 1541–1568, doi: https://doi.org/10.1175/WAF-D-16-0178.1

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014,0155:HTFENP.2.0.CO;2.

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Q.J.R. Meteorol. Soc.*, **131**, 3269-3289. https://doi.org/10.1256/qj.05.135.

Hu, J., N. Yussouf, D. D. Turner, T. A. Jones, and X. Wang, 2019: Impact of ground-based remote sensing boundary layer observations on short-term probabilistic forecast of a tornadic supercell event. Mon. Wea. Rev., 34, 1453–1476. https://doi.org/10.1175/WAF-D-18-0200.1.

Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-based evaluation of the impact of horizontal grid spacing on convection-allowing forecasts. *Mon. Wea. Rev.*, **141**, 3413–3425, https://doi.org/10.1175/MWR-D-13-00027.1.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

——, P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith, 2018: Comparison of cloud microphysics schemes in a Warn-on-Forecast System using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, https://doi.org/10.1175/WAF-D-18-0112.1.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

——, P.S. Skinner, N. Yussouf, K. Knopfmeier, A.E. Reinhart, and D.C. Dowell, 2019: Forecasting high-impact weather in landfalling tropical cyclones using a Warn-on-Forecast system. *Bull. Amer. Meteor. Soc.*, **100**, 1405–1417. https://doi.org/10.1175/BAMS-D-18-0203.1.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, https://doi.org/10.1175/WAF2007106.1.

——, P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research Scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806, https://doi.org/10.1175/BAMS-84-12-1797.

Kerr, C.A., B. C. Matilla, Y. Wang, D. R. Stratman, T. A. Jones, and N. Yussouf, 2022: Results from a pseudo-realtime next-generation 1-km Warn-on-Forecast system prototype. *Wea. Forecasting, submitted.*

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D.M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

——, C. K. Potvin, P. S. Skinner, and A. E. Reinhart, 2021: The Vice and virtue of increased horizontal resolution in ensemble forecasts of tornadic thunderstorms in low-CAPE, high-shear environments. *Mon. Wea. Rev.*, **149**, 921–944, https://doi.org/10.1175/MWR-D-20-0281.1.

Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2017: Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. *Wea. Forecasting*, **32**, 1403–1421, https://doi.org/10.1175/WAF-D-16-0200.1.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Miller, W., C. K. Potvin, M. L. Flora, B. T. Gallo, L. J. Wicker, T. A. Jones, P. S. Skinner, B. Matilla, and K. H. Knopfmeier, 2021: Exploring the usefulness of downscaling free forecasts from the Warn-on-Forecast System. *Weather and Forecasting*, **37**, 181-203, https://doi.org/10.1175/WAF-D-21-0079.1.

Nachamkin, J. E., and J. Schmidt, 2015: Applying a neighborhood fractions sampling approach as a diagnostic tool. *Mon. Wea. Rev.*, **143**, 4736–4749, doi:10.1175/MWR-D-14-00411.1.

Niu, G. Y., et al., 2011: The community Noah land surface model with multi-parameterization options (Noah- MP): 1. Model description and evaluation with local- scale measurements." *Journal of Geophysical Research: Atmospheres*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-on-Forecast. *Mon. Wea. Rev.*, **143**, 2998–3024, https://doi.org/10.1175/MWR-D-14-00416.1.

——, E. M. Murillo, M. L. Flora, and D. M. Wheatley, 2017: Sensitivity of supercell simulations to initial-condition resolution. *J. Atmos. Sci.*, **74**, 5–26, https://doi.org/10.1175/JAS-D-16-0098.1.

Roberts, B., B. T. Gallo, I. L. Jirak, and A. J. Clark, 2019: The High Resolution Ensemble Forecast (HREF) system: Applications and performance for forecasting convective storms. 2019 AGU Fall Meeting, San Francisco, CA, Amer. Geophys. Union, Abstract A31O-2797.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Schumacher, R. S., 2015: Resolution dependence of initiation and upscale growth of deep convection in convection-allowing forecasts of the 31 May–1 June 2013 supercell and MCS. *Mon. Wea. Rev.*, **143**, 4331–4354, https://doi.org/10.1175/MWR-D-15-0179.1.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF Model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372, https://doi.org/10.1175/2009MWR2924.1.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

——, 2014: Reproducing the September 2013 record-breaking rainfall over the Colorado Front Range with high-resolution WRF forecasts. *Wea. Forecasting*, **29**, 393–402, https://doi.org/10.1175/WAF-D-13-00136.1.

——, G. S. Romine, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. *Mon. Wea. Rev.*, **145**, 2943–2969, https://doi.org/10.1175/MWR-D-16-0410.1.

——, and R. A. Sobash, 2019: Revisiting sensitivity to horizontal grid spacing in convection-allowing models over the central and eastern United States. *Mon. Wea. Rev.*, **147**, 4411–4435, https://doi.org/10.1175/MWR-D-19-0115.1.

Shao, H., and Coauthors, 2016: Bridging Research to Operations Transitions: Status and Plans of Community GSI. *Bulletin of the American Meteorological Society*, **97**, 1427–1440, https://doi.org/10.1175/BAMS-D-13-00245.1.

Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., http://dx.doi.org/10.5065/D68S4MVH.

Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast system. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Sobash, R. A., C. S. Schwartz, G. S. Romine, and M. L. Weisman, 2019: Next-day prediction of tornadoes using convection-allowing models with 1-km horizontal grid spacing. *Wea. Forecasting*, **34**, 1117–1135, https://doi.org/10.1175/WAF-D-19-0044.1.

Sobash, R. A., and L. J. Wicker, 2015: On the impact of additive noise in storm-scale EnKF experiments. *Mon. Wea. Rev.*, **143**, 3067–3086, https://doi.org/10.1175/MWR-D-14-00323.1.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

VandenBerg, M. A., M. C. Coniglio, and A. J. Clark, 2014: Comparison of next-day convection-allowing forecasts of storm motion on 1- and 4-km grids. *Wea. Forecasting*, **29**, 878–893, https://doi.org/10.1175/WAF-D-14-00011.1.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Wang, Y., and X. Wang, 2020: Prediction of tornado-like vortex (TLV) embedded in the 8 May 2003 Oklahoma City tornadic supercell initialized from the subkilometer grid spacing analysis produced by the dual-resolution GSI-based EnVar data assimilation system. *Mon. Wea. Rev.*, **148**, 2909–2934, https://doi.org/10.1175/MWR-D-19-0179.1.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman Filter Analyses and Forecasts of a Severe Mesoscale Convective System Using Different Choices of Microphysics Schemes. *Mon. Wea. Rev.*, **142**, 3243–3263, https://doi.org/10.1175/MWR-D-13-00260.1.

——, K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar Data Experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Wilson, K. A., P. L. Heinselman, P. S. Skinner, J. J. Choate, and K. E. Klockow-McClain, 2019a: Meteorologists' Interpretations of Storm-Scale Ensemble-Based Forecast Guidance. *Wea. Climate Soc.*, **11**, no. 2 (April): 337-354. https://doi.org/10.1175/WCAS-D-18-0084.1.

Wilson, K. A., J. J. Choate, A. J. Clark, B. T. Gallo, P. L. Heinselman, K. H. Knopfmeier, B. Roberts, P. S. Skinner, and N. Yussouf, 2019b: Exploring Applications of Storm-Scale Probabilistic Warn-on-Forecast Guidance in Weather Forecasting, *International Conference on Human-Computer Interaction*, pp. 557–572. Springer, Cham., https://doi.org/10.1007/978-3-030-21565-1_39.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, and K. K. Droegemeier, 2013: Prediction of convective storms at convection-resolving 1-km resolution over continental United States with radar data assimilation: An example case of 26May 2008 and precipitation forecasts from spring 2009. *Adv. Meteor.*, 2013, https://doi.org/10.1155/2013/259052.

Yang, Z.-L., G.-Y. Niu, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, K. Manning, D. Niyogi, M. Tewari, and Y. Xia, 2011: The Community Noah Land Surface Model with Multi-Parameterization Options (Noah-MP): 2. Evaluation over Global River Basins. *J. Geophys. Res.*, **116**, D12, https://doi.org/10.1029/2010JD015140.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.

Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, https://doi.org/10.1175/MWR-D-12-00237.1.

——, D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, https://doi.org/10.1175/MWR-D-14-00268.1.

——, J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Wea. Forecasting*, **31**, 957–983, https://doi.org/10.1175/WAF-D-15-0160.1.

——, and K. H. Knopfmeier, 2019: Application of Warn-on-Forecast system for flash-flood producing heavy convective rainfall events. *Quart. J. Roy. Meteor. Soc.*, **145**, 2385–2403, https://doi.org/10.1002/qj.3568.

——, T. A. Jones, and P. S. Skinner, 2020: Probabilistic high-impact rainfall forecasts from landfalling tropical cyclones usingWarn-on-Forecast system. *Quart. J. Roy. Meteor. Soc.*, **146**, 2050–2065. https://doi.org/10.1002/qj.3779.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0094.1.